Experimental design and statistics for Marine Ecology

Part 3

STATISTICAL ANALYSIS AND DESIGN OF ECOLOGICAL EXPERIMENTS - ANALYSIS OF VARIANCE

Written by P Jonsson 1995, modified by M Lindegarth, C Andre 2001 and by P Jonsson 2008 & 2009

The design & analysis of more complex scientific studies

In Part 1 we used the *t*-test to test the probability that two samples come from the same true population. Examples include tests of field observations, e.g. comparison of mussel lengths between two locations, or in manipulative experiments, e.g. testing a model of the effect of two diets on growth rate. In Part 2 we extended this to tests of many samples using linear regression. However, linear regression requires that there is a simple relationship (linear) between the measured dependent variable and the different levels of the factor. Real scientific models are usually more complex and require more than two samples. Here are three examples:

1. There are more than two samples that are tested against the H_0 of no difference. In Fig. 1A we are comparing the concentration of a defensive chemical compound in macroalgae among the locations. In other words the factor *Locations* has three levels, and there are three true statistical populations.
2. There may be several hierarchical spatial factors like in Fig. 1B where the hypothesis is that processes acting on different scales, e.g. salinity gradients and wave exposure, affect the abundance of a polychaete. The design in Fig. 1B allows for a test if and how much of the variation in polychaete abundance that is explained by the factors *Bay*, *Site* and among replicates (the residual variance).
3. An experiment is designed as a combination of factors. In Fig. 1C a test is performed of how the induction of a chemical defence compound responds to the factor *Grazer* (presence/absence) and to the factor *Nutrients* (low/high). One model makes the prediction that induction of the compound occurs when a grazer is present but only when nutrients are scarce. When the response to a factor depends on the levels of another factor this is called an interaction between factors and needs more advanced statistical methods.

To test the hypotheses in the more complex designs in Fig. 1 we need more advanced statistical methods. One powerful method is the analysis of variance (ANOVA) and we will devote this Part 3 entirely to learn the technique of ANOVA. We have already met ANOVA in a simple form in Part 2 where it was used in regression analysis. Here we will extend ANOVA to many levels and multiple factors.

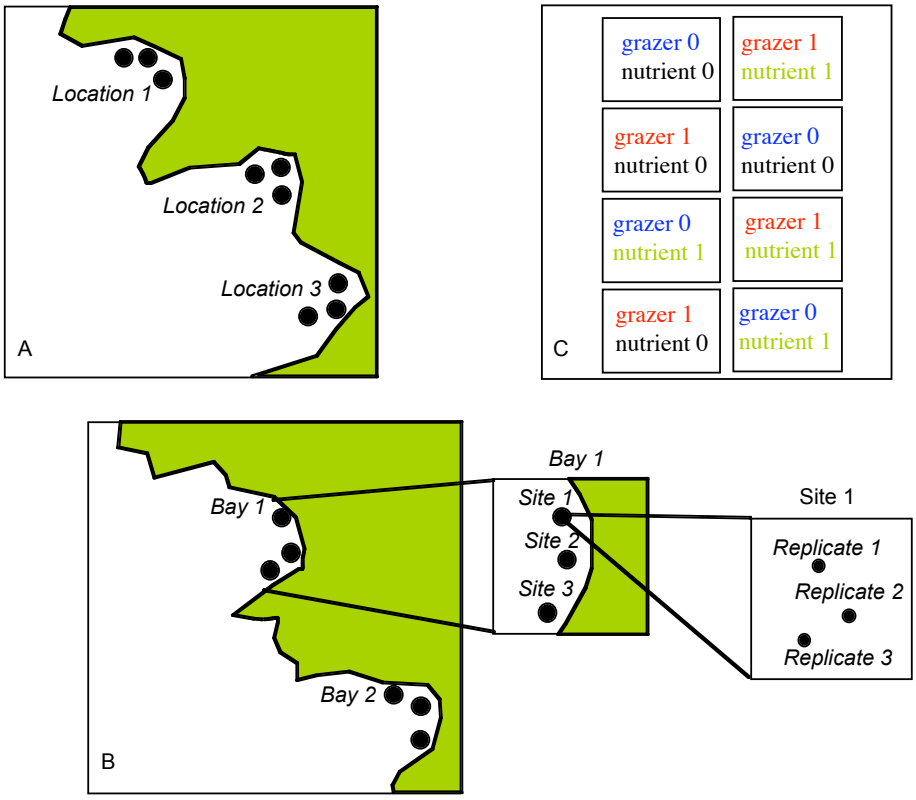

Fig. 1. Examples of more complex observations and experiments where ANOVA is suitable.

One-factor ANOVA

Imagine that we want to test if 3 species of copepods differ in their ability to induce a toxin in a dinoflagellate. The factor *Copepod* has 3 levels and we could test the H_0 that the species do not differ by performing 3 different *t*-tests. There are, however, two problems:

1. With increasing number of factor levels the number of pairwise tests rapidly becomes very many (Table 1).
2. With increasing number of pairwise tests of the same samples we will increase the type 1 error (Table 1).

Table 1. More levels increase the number of tests and the type 1 error.

Number of levels	*Number of tests (k)*	*Type 1 error, 1-(1-a)^k*
2	1	0.05
3	3	0.14
5	10	0.40
10	45	0.90
15	105	0.995

The problem with tests of factors including more than 2 levels was solved by the outstanding statistician and evolutionary biologist Ronald Fisher in the 1930ies. Fisher

started with a number of samples and proposed the H_0 that $\mu_1 = \mu_2 = \mu_i = ... = \mu_a$ (Fig. 2). If H_0 is false the means are different, e.g. due to an experimental treatment effect A_1, A_2, $A_i ... A_a$ (Fig. 2). To remove any effect of the sign of these treatments they can be squared and H_0 can be re-formulated to state that:

$$H_0 = \sum_{i=1}^{a} A_i^2 = 0$$

and the alternative hypothesis of an effect from the treatments is:

$$H_A = \sum_{i=1}^{a} A_i^2 > 0$$

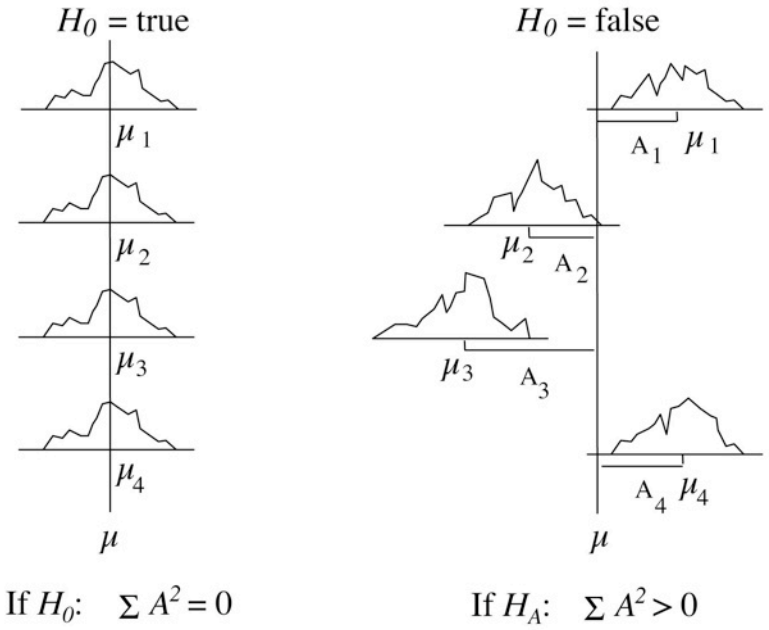

Fig. 2. Schematic drawing of population distributions and treatment effects (A) when H_0 is true (left) and false (right), respectively.

A really smart feature of formulating H_0 in this way is that the test is one-sided. Any alternative hypothesis will make $\sum A_i^2$ larger than for H_0. The trick is now to understand how $\sum A_i^2$ can be estimated from a number of samples and how we know when this sum of treatment effects are sufficiently high to reject H_0. In Part 2 the concept of a linear model was introduced. We can write a linear model to account for all the variation among and within samples as:

$$X_{ij} = \mu + A_i + e_{ij}$$

As before X_{ij} is the particular value of object j in treatment i and μ is the true mean, A_i are the different levels of treatment effects, and e_{ij} are the deviations of each individual

observation (measurement) not explained by the mean or the treatment effect. Fig 3 shows how the total variation within and among samples can be partitioned into three different parts by measuring the variation as the Sum of Squared deviations (*SS*).

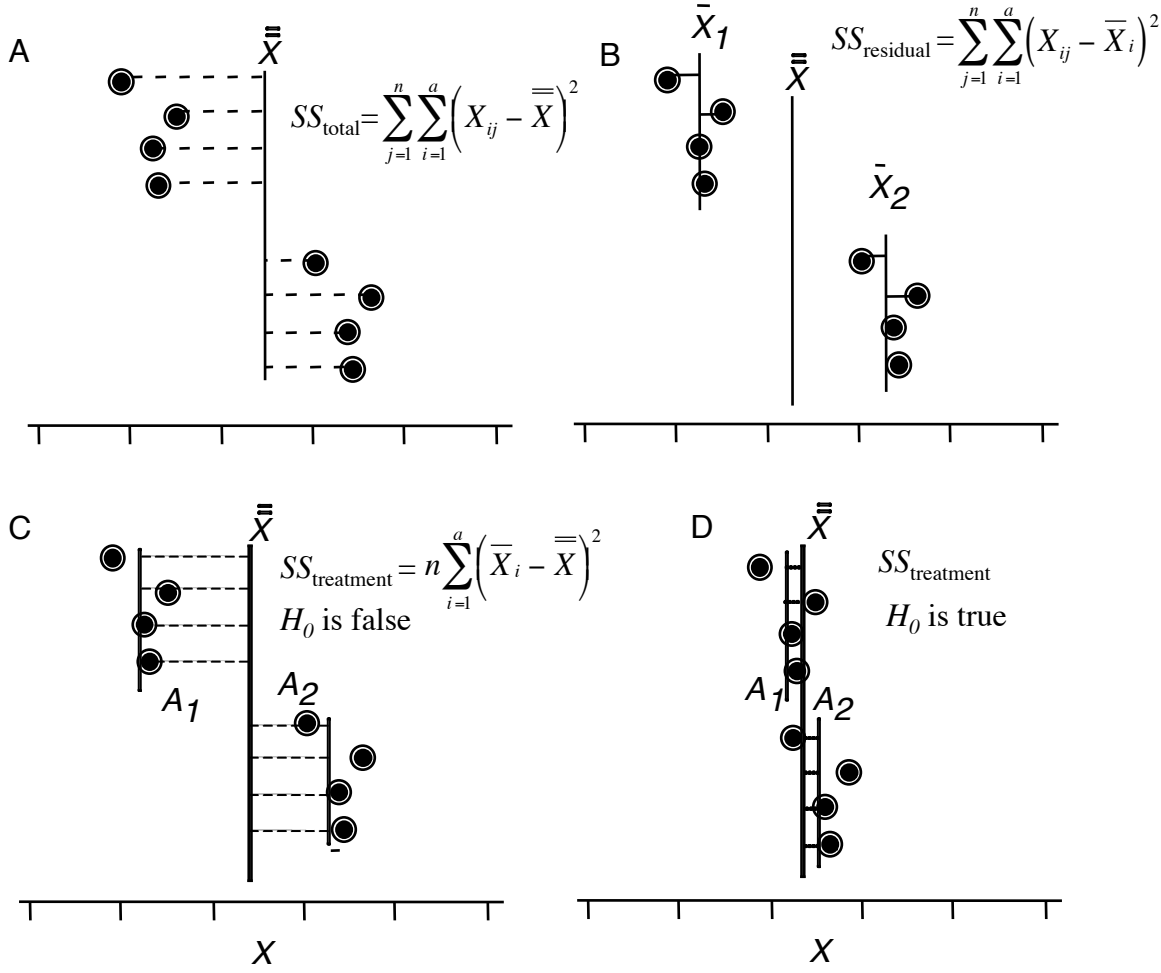

Fig. 3. Drawing showing the construction of the different Sums of Squares (*SS*), see text below.

We can walk through Fig. 3 in steps:

1. Fig. 3A illustrates the <u>total variation</u> measured as the *SS* between each observation and the overall mean ($\overline{\overline{X}}$). The *SS* is calculated as $\sum_{j=1}^{n}\sum_{i=1}^{a}\left(X_{ij} - \overline{\overline{X}}\right)^2$.
2. The total variation can be partitioned into two parts. Fig. 3B shows the variation of each observation (the circles in Fig. 3) to the means ($\overline{X}_i$) of the two levels of Factor *A*. This variation is called the <u>residual *SS*</u> ($SS_{Residual}$) and is calculated as $\sum_{j=1}^{n}\sum_{i=1}^{a}\left(X_{ij} - \overline{X}_i\right)^2$. The $SS_{Residual}$ represents the variation that is <u>not explained</u> by the factor *A*.

3. The second contribution to the total variation is shown in Fig. 3C and is the part explained by the factor *A*. The $SS_{treatment}$ for factor *A* is calculated as $n\sum_{i=1}^{a}\left(\overline{X}_i - \overline{\overline{X}}\right)^2$. Note that the $SS_{treatment}$ for factor *A* is multiplied by the sample size *n*. Why? One simple explanation is that each *SS* should consist of the same number of deviations to be able to compare them. For each factor level the deviation $(\overline{X}_i - \overline{\overline{X}})^2$ should be taken for each observation and then summed to get the $SS_{treatment}$. Of course there will then be *n* identical $(\overline{X}_i - \overline{\overline{X}})^2$ so we can write the sum as $n^* \Sigma(\overline{X}_i - \overline{\overline{X}})^2$. That this really makes statistical sense can be seen if we think about what the different *SS* estimate. If we assume that the factor *A* has no effect (no treatment effect) and H_0 is true then ΣA_i^2 is expected to be 0. This is the case in Fig. 3D. $SS_{treatment}$ for factor *A* will still be different from 0 in a given sample due to non-representativity and this variation is caused by the fact that the individual observations differ. What the $SS_{treatment}$ really estimates is the variation of the mean for all individual observations. Remember the Central Limit Theorem and that the variance of a mean is s^2/n, i.e. the sample variance divided by sample size. So if $\Sigma(\overline{X}_i - \overline{\overline{X}})^2$ is multiplied by *n* (sample size) and there is no effect of Factor *A* the *SS* in Fig 3D should estimate the same variation as the $SS_{residual}$. This is exactly what we want $SS_{treatment}$ for factor *A* to do, i.e. when there is no effect of factor *A* its *SS* only estimates the $SS_{residual}$.

The final step is to construct a table for our one-factor ANOVA. We did this already in Part 2 for the regression analysis and the table for the 1-factor ANOVA is very similar. Table 2 shows how we collect our information about our attempt to partition the variation.

Table 2. One-factor analysis of variance.

Source of variation	***SS***	***df***	***MS***	***MS* estimates**
Factor A	$n\sum_{i=1}^{a}\left(\overline{X}_i - \overline{\overline{X}}\right)^2$	*a*-1	$\frac{SS_{treatment}}{(a-1)}$	$\sigma^2_e + \frac{n\sum A_i^2}{(a-1)}$
Residual (*within groups*)	$\sum_{j=1}^{n}\sum_{i=1}^{a}\left(X_{ij} - \overline{X}_i\right)^2$	*an-a*	$\frac{SS_{residual}}{(an-n)}$	σ^2_e
Total	$\sum_{j=1}^{n}\sum_{i=1}^{a}\left(X_{ij} - \overline{\overline{X}}\right)^2$	*an*-1	$\frac{SS_{total}}{(an-1)}$	

The first column has the different sources of variation we already recognize from the linear model and Fig. 3. The second column shows the *SS*. Note again that the $\Sigma\Sigma(\overline{X}_i - \overline{\overline{X}})^2$ for $SS_{treatment}$ can be written as $\boldsymbol{n}^* \Sigma (\overline{X}_i - \overline{\overline{X}})^2$.

Next step is to form a measure of variation that is independent of sample size. Note that the *SS* will increase as we increase the sample size of an experiment; when *n* increases so do the number of squared deviations. This is of course unfortunate because we want to obtain a statistic that is general across all types of studies regardless of the size of an experiment. To achieve this the *SS* are divided by the degrees of freedom (*df*) in the third column. We then get what is called mean squares (*MS*) and are shown in the fourth column. A *MS* is similar to a variance since we divide a *SS* with the number of independent deviations, i.e. *df*. As before the number of *df* is found as the number of squared deviations minus the number of parameters (means) we need for the squared deviations. For example for a factor with *a* levels, the $SS_{treatment}$ consists of *a* deviations and we need to calculate the overall mean so this gives *a*-1 *df*, and the $SS_{residual}$ has *a*n* deviations and we need *a* means of the different factor levels giving a *df* of *a*n-a*. The final step (often the most tricky!) is to find out what the different *MS* estimate, this is written in the fifth column. *MS* is similar to a variance, but we will see that an *MS* may estimate the sum of many variances. As shown in Fig. 3D the $SS_{treatment}$ and $MS_{treatment}$ will under H_0, of no treatment effect, estimate the same variation as $SS_{residual}$ and $MS_{residual}$.

We are now ready to find a statistical test for a possible effect of factor *A*. As is seen in Table 2 the $MS_{treatment}$ will estimate the same variance as the $MS_{residual}$ if H_0 is true, i.e. if there is no treatment effect from our manipulation of factor *A*. As you may guess the proper statistic to test this H_0 is the *F*-ratio, where $F = MS_{treatment} / MS_{residual}$. Under H_0 the expected *F* is close to 1, and for alternative hypotheses *F* will increase. The distribution of *F* under H_0 depends on the *df* both for the *MS* in the numerator and the denominator. The critical *F* for rejection of H_0 may be found in statistical tables (or as a formula in Excel).

Assumptions for ANOVA

1. Samples come from normal distributions. As pointed out before the probability of an *F*-ratio will only be exactly valid if samples come from a normal distribution. But this assumption is not very critical for ANOVA because we compare means and due to the Central Limit Theorem these are almost always close to normally distributed. In fact, ANOVA is very robust (not sensitive) even when the true populations are far from normally distributed.
2. Variances are homogeneous (homoscedasticity). We have already said that this assumption is more important than the assumption about normal distributions for most

parametric techniques. Especially, if one or a few variances are much larger than in the other groups this may have serious effects on the type I error. Usually, heterogeneous variances lead to an excessive type I error and we will reject H_0 more often than our calculated probability indicates. Before performing an ANOVA we need to inspect the data to identify suspect measurements, and will also formally test the H_0 that variances are coming from the same distribution, i.e. they are homogeneous. There are many tests but we recommend Cochran's test where the *C*-statistic is easy to calculate:

$$C = \frac{\max\left(s_i^2\right)}{\sum s_i^2}$$

where s_i^2 are the variances of all the treatment groups and *max* represents the largest of the variances. The probability distribution of *C* can be found in statistical tables and the *C*-statistic depends on the number of variances compared and the *df* of each variance (*n*-1). An example is where the factor *A* has 4 levels with 5 replicate observations for each level, i.e. sample size is 5. The variances are 5.3, 6.4, 7.8 and 15. The sum of these variances are 34.5 and the maximum variance is 15 so the statistic *C* is 0.43 with 4 variances and 4 *df*. A table shows that the critical value to reject H_0 of homogeneous variances in this case is 0.63 so with our lower value of 0.43 we consider the variances homogeneous and we can proceed with the ANOVA.

3. Independent observations. As before it is of paramount importance that our observations, or replicates, are independently sampled. Failure to meet this assumptions have unpredictable consequences and any statistical analysis will be compromised. This is probably the most common error in scientific studies. Here are some examples of when observations are not independent:
   i) Within a level: Crabs within the same aquarium stimulate each other to release a pheromone, which leads to a decrease in the variation among crabs.
   ii) Between levels: In a field experiment plots are prepared by removing all grazers. Some randomly selected plots are assigned as controls and grazers are returned. Unfortunately, the plots with and without grazers are too close so many of the grazers migrating into the empty plots come from the control plots. Clearly, the observations between treatments are not independent and in this case the effect of grazing will be underestimated.
4. Balanced samples. The final assumption is that, as far as possible, any study should be what is called balanced. This means that there should be an equal number of observations or replicates in all samples or groups. Unbalanced designs may lead to several errors, especially when we extend our ANOVAs to include more than one factor (see below). Equally important is that balanced designs are more robust against heterogeneous variances and non-normality. Although we plan a balanced design we may lose observations, e.g. an organism died or we dropped a test tube. Possible

solutions may include that we randomly remove an observation from the other groups to maintain the balanced design. We may then compare the analysis of the reduced but balanced data set to an analysis of the unbalanced data set. If conclusions are different it may be necessary to seek statistical expertise. If there are very few replicates in each group the removal of observations may lead to an increased type II error. We may here replace the missing observation by the mean in that group, and in the final analysis remove one degree of freedom in the denominator of the *F*-test.

Multiple comparisons or a posteriori *tests*

An ANOVA is rarely the endpoint for our test of hypotheses. As we have seen, the ANOVA only tests the general hypothesis that a factor explains a significant part of the variation observed. The ANOVA does not tell anything about what levels in a factor are responsible for the treatment effect. For example, a compound extracted from a sponge is suspected to inhibit settlement of fouling organisms, e.g barnacles. Barnacle larvae are added to four concentrations of the compound and one control, and the proportions of settled larvae in 6 replicate dishes are recorded (there are 20 larvae per dish). In a one-factor ANOVA we get a sufficiently large F to reject the H_0 of no effect. The hypothesis that the compound has an antifouling effect is supported. However, the ANOVA says nothing about which of the concentrations had an effect. If there are more specific hypotheses about the different levels in a factor this requires more detailed tests. These tests are called multicomparison tests, and one type is called *a posteriori tests*. They are related to pairwise *t*-tests but include some insurance against the excessive type I error that we saw in Table 1. A multicomparison test is only done if the ANOVA rejects the H_0. There are several multicomparison tests and one commonly used is the Student-Newman-Keuls test (SNK). Briefly, this test starts with the ranking of all group means, i.e. the means of all levels. First a *SE* is calculated from the *MS* in the denominator in the *F*-ratio (often termed MS_{error}) as:

$$SE = \sqrt{\frac{MS_{error}}{n}}$$

where n is the number of replicates in each group. This *SE* has the same *df* as the MS_{error}. Secondly, we begin to test the means being most different (H_0 of no difference) with the statistic Q as:

$$Q = \frac{\overline{X}_i - \overline{X}_j}{SE}$$

The distribution of Q (or critical values of the distribution) under H_0 can be found in statistical tables where Q depends on the number of means between the tested pair of means and the *df* of *SE*. In the ANOVA of the test of our sponge compound we got an $F_{4,25}=MS_{treatment} / MS_{residual}=4755/100=47.5$. This F is much larger than the critical 2.6 and

we reject H_0. We then proceed with an SNK-test of the means. The 5 means are ranked as: 70%, 60%, 45%, 15% and 5% so we start to test 70% against 5% which has a "distance" of 5 means. The *SE* is $\sqrt{100/6} = 4.08$, and the Q=(70-5)/4.08=16.9. The critical Q with α=0.05 with 25 *df* and a distance of 5 means is found in a table as 4.17. Our Q of 16.9 clearly exceeds this and we consider the means 70 and 5% as significantly different. We then proceed with the next greatest difference in means, here 60 and 5% or 70 and 15%. When a difference is not found to be significantly different we stop our multicomparison. The idea with this procedure is to perform a minimum of comparisons to minimize the type I error. There is an alternative way to *a posteriori* multicomparison tests called *a priori* tests. Read more about this on page 20.

ANOVA hierarchical factors - Nested ANOVA

We will now begin to include more than one factor in an ANOVA and we start with what is called nested (or hierarchical factors). To illustrate the logic behind a nested ANOVA we start with an example. We want to investigate if the number of grazers (e.g. the isopod *Idotea viridis*) on bladder wrack (*Fucus vesiculosus*) differs between sheltered and wave exposed shores. First we identify possible exposed and sheltered sites and randomly select two of each. At these sites we sample a total of 20 plants according to the map in Fig. 4. Table 3 shows the results of the study with all the means.

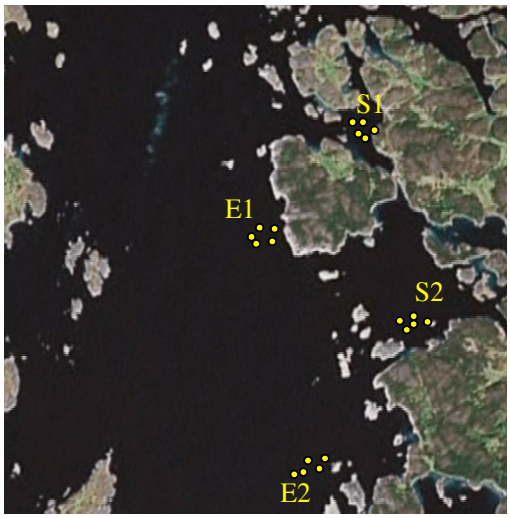

Fig. 4. Map showing the location of 20 samples of bladder wrack from sheltered and exposed sites.

Table 3. Number of isopods on bladder wrack for 20 plants sampled from both exposed and sheltered shores.

Exposed 1	*Exposed* 2	*Sheltered* 1	*Sheltered* 2
10	15	27	38
23	38	18	49
45	18	35	53
12	25	19	37
21	20	24	48
$\overline{X}_{E1}$=22.2	$\overline{X}_{E2}$=23.2	$\overline{X}_{S1}$=24.6	$\overline{X}_{S2}$=45
	$\overline{X}_{E}$=22.7		$\overline{X}_{S}$=34.8
			$\overline{\overline{X}}_{total}$=28.7

We first analyze this data set according to a one-factor ANOVA with *Exposure* as the factor with two levels (exposed/sheltered). Table 4 shows the result of the ANOVA with an *F*-ratio for 1 and 18 *df* equal to 5.2 and we reject the H_0 of no difference and conclude that grazers indeed differ between exposed and sheltered sites. However, there is something wrong with this analysis. It is clear from the map in Fig. 4 that the 20 sampled plants

Table 4. Results of ANOVA of the number of isopod grazers present on bladder wrack at exposed and sheltered sites.

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Exposure	732	1	732	$F_{1,18}$=5.2	0.035
Residual	2535	18	141		
Total	3267	19			

are not independent. They appear in groups of 5 and all the 10 plants representing exposed and sheltered shores were not randomly distributed. For the test of the main hypothesis that there is a difference between exposed and sheltered shores there are only 2 independent replicates, *E1* and *E2* for the exposed level and *S1* and *S2* for the sheltered level.

A possible way to correct this mistake would be to calculate the means for each of the 4 sites. Now the ANOVA will look like in Table 5. Note that the degrees of freedom for the *F*-test decreased dramatically reflecting that we only have 2 independent replicates. Accordingly, the probability of H_0 being true is high and in contrast to the ANOVA in Table 4 we retain H_0.

Table 5. ANOVA of the same data as in Table 4 but now each site is represented by an independent mean.

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Exposure	146	1	146	$F_{1,2}$=1.4	0.36
Residual	209	2	104		
Total	355	3			

There is, however, another way of analyzing the data collected as in Fig. 4. We can explicitly include the spatial dependence of the 5 plants sampled at each site. We do this by adding a second factor, *Site*, to our linear model:

$$X_{ijk} = \mu + Exposure_i + Site\ (Exposure)_{j(i)} + e_{ijk}$$

To indicate that Site is nested under *Exposure* we use brackets: *Site(Exposure)*. The logic behind nested factors is a bit difficult to see at first but be patient! If we look at the map again in Fig. 5 we now think like this: there are two

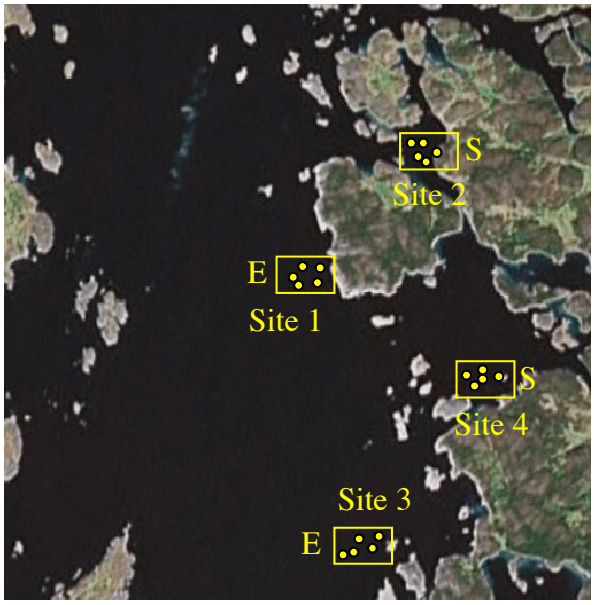

Fig 5. A map of the same samples as in Fig. 4 but where the grouping in Factor *Site* is shown.

randomly selected *Sites* for each level of *Exposure* (exposed/sheltered). Each level of the factor *Site* only represents a different place. And each level of *Site* is found only under one level of *Exposure*, i.e. the factor *Site* is nested under the factor *Exposure*. Under the factor *Site* there are 5 plants in each level. The ANOVA of this nested design will look like in Table 6. The numbers are partially the same as in the naïve ANOVA in Table 4, but the $SS_{Residual}$ in Table 4 is now partitioned between the nested factor *Site* and a new $SS_{Residual}$. The SS_{Site} of course contains the variation explained by being at different sites regardless of exposure. The important thing to notice in Table 6 is that the relevant *F*-test for the H_0 that

Exposure does not have an effect is $MS_{Exposure} / MS_{Site}$. We easily see this when we look at what the *MS* estimates in Table 6. For the test for *Exposure* we should identify a denominator *MS* (also called error term) that contains all the variance components except the one we are testing. Clearly, MS_{Site} is here the correct error term. We also see by testing Exposure over the nested factor Site that we get the same F=1.4 as when we tested only the means in Table 5.

Table 6. ANOVA of the same data as in Table 4 but analyzed according to a nested ANOVA design. Note that the *A* and *B* after the factors are just to make the notation of what *MS* estimates shorter.

Source of Variation	*SS*	*df*	*MS*	*F*	*MS estimates*	*P*
Exposure (A)	732	1	732	1.4	$\sigma^2_e + n\sigma^2_{B(A)} + \frac{nb\sum A_i^2}{a-1}$	0.35
Site (*Exposure*) (*B*)	1043	2	521	5.6	$\sigma^2_e + n\sigma^2_{B(A)}$	0.014
Residual	1492	16	93		σ^2_e	
Total	3267	19				

But we also get some more information. We see that we have to reject the H_0 for factor *Site* and there are clearly differences among sites that are not explained by exposure. It was really this effect we incorrectly interpreted as an effect of exposure in Table 4. By including nested factors in this hierarchical way we can gain valuable information about variation in the system we study. With the information in Table 6 it is, e.g. possible to calculate the optimum compromise between number of sites visited and the number of plants analyzed (see below).

The next example of a nested ANOVA is about a manipulative experiment where we want to test the model that growth rate of a fish depends on the type of diet. The experiment is designed with 3 aquaria receiving Diet 1 and 3 aquaria with Diet 2. In each aquarium there are 5 fish which are measured before and after the experiment. The dependent variable that we will analyze using an ANOVA is the weight increase. The H_0 is that there is no difference in weight increase between Diet 1 and 2. The variation in weight increase is analyzed according to the linear model:

$$X_{ijk} = \mu + Diet_i + Aquarium(Diet)_{j(i)} + e_{ijk}$$

The linear model shows that we now have learnt about nested factors and realized that the 5 individual fish in each aquarium are dependent observations; the independent experimental units are the aquaria. To represent the actual design in the ANOVA we identify the factor *Diet* and the factor *Aquarium(Diet)* nested under *Diet*. The correct ANOVA is schematically shown in Table 7. Again we see that the correct *F*-test for the

Table 7. ANOVA of two factors where one (*Aquarium*) is nested under the other (*Diet*).

Source of Variation	*SS*	*df*	*MS*	*F*	*MS estimates*
Diet		a-1			$\sigma^2_e + n\sigma^2_{B(A)} + \frac{nb\sum A_i^2}{a-1}$
Aquarium(Diet)		ab-a			$\sigma^2_e + n\sigma^2_{B(A)}$
Residual		abn-ab			σ^2_e
Total		abn-1			

factor *Diet* is formed by $F=MS_{Diet} / MS_{Aquarium}$. A significant effect of *Aquarium* means that different aquaria, for some reason, resulted in different growth rates despite receiving the same diet. Maybe some aquaria were placed closer to the door in the constant temperature room and were disturbed more severely when people walked in and out. Or maybe the fish were not randomly distributed among the aquaria. If the fish were caught in a trap net from a storage tank and aquaria were filled one by one it is likely that the first few aquaria received the slowest and perhaps inferior individuals (how could this be avoided?).

In Figure 6 the logic behind nested ANOVA is shown in graphic form. It is very, very common that scientific studies contain designs where observations are dependent without including this in the statistical analysis. This has the effect that the test of H_0 has too many *df* and a type I error is committed (as seen in Table 4). The problem is particularly common when samples are collected from different localities in the field (as in our example above) or when the experimental units are big. Many experiments feature large containers that can hold many m^3 of water, so called mesocosms. It is easy to forget that these are still the independent experimental unit. In too many studies there is just one mesocosm for each treatment level, and then a number of dependent samples from each mesocosm. All that is logically tested in such a design is if two containers differ in some aspect, and often they do. However it is illogical to conclude that this is linked to the treatment. Instead, at least two mesocosms per treatment level are required and the dependent samples are included as a nested factor.

Pooling of nested effects

If the factor *Site* in Table 6 had turned out to be non-significant this would have indicated that there was no large difference among the different sites. Many statisticians (but not all!) recommend that it is then possible to view the different plants as independent observations to estimate the residual variation. Often it is recommended that the probability that H_0 is true should be greater than 0.25. If this had been the case in Table 6 we can now pool together the variation explained by the factor *Site* and the unexplained variation in the

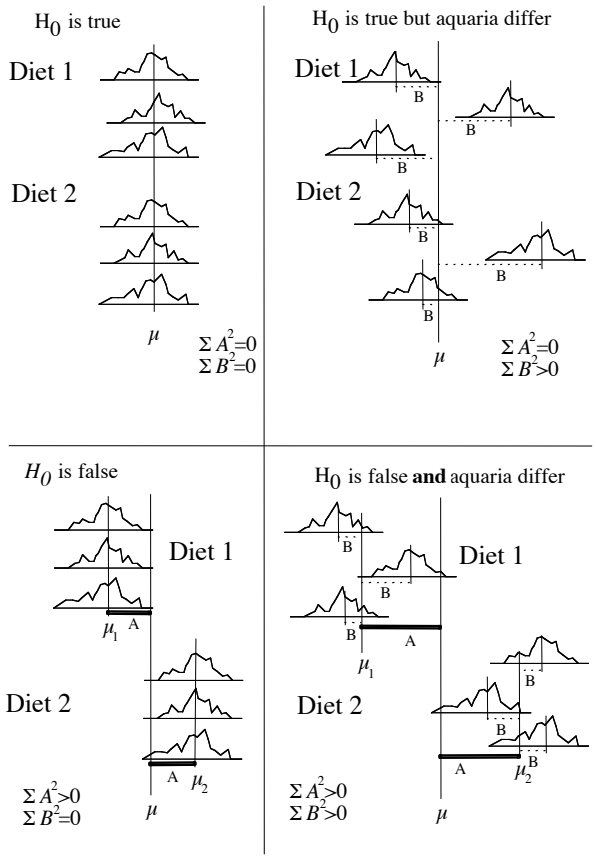

Fig. 6. Schematic drawing of the logic behind nested ANOVA. All possible cases for the tests of the two H_0 are shown.

Residual and form a new $MS_{Residual}$ called MS_{Pooled}. We do this by adding the *SS* and divide by the sum of the *df*:

$MS_{Pooled} = (1492+1043)/(16+2) = 141$

The *F*-ratio is now 5.2 and we arrive at the same test as in Table 4, BUT now we have formally tested if sites differ.

Why include nested factors?

There are two major reasons why we decide to include nested factors in our designs.

1. We may have a limited number of aquaria but instead of adding just one fish per aquarium we add several fish that give the opportunity to explore if the experimental units (the aquaria) differ much. Much variation among aquaria indicates that there is one (or several) unknown factors that affect the experiment. Maybe it is possible to identify these and design a stronger experiment.
2. Generally resources (usually time & money) are limited for any study. In the experiment above with the analysis of number of grazers on plants of *Fucus vesiculosus*, there is one cost to travel to a new site and one cost to analyze every plant. In a nested design it is possible to make a cost-benefit analysis to decide for a

given amount of resources what is the optimum number of sites and plants to maximize the statistical power. It is possible to show that for the total available resources C (e.g money) the optimum number of sites (b) per treatment level and the number of plants per site (n) are:

$$b = \frac{C}{C_b + nC_n}, n = \sqrt{\frac{C_b \sigma_e^2}{C_n \sigma_{B(A)}^2}}$$

where C_b is the cost of visiting a new site, C_n is the cost per plant, $\sigma^2_{B(A)}$ is the variation explained by *Site* and σ^2_e is the residual variation (between plants within a site). If we use the equation above and the *MS* in Table 6 we can calculate the variation $n\sigma^2_{B(A)}$ (a sort of variance) explained by *Site* by using the *MS* and what they estimate. First the σ^2_e = 93 and:

$$\sigma^2_{B(A)} = (MS_{Site} - MS_{Residual})/n = 86$$

Imagine that we now have a total budget of 50000 SEK. To visit a new site costs 3000 SEK in ship time and to analyze one plant of *Fucus* costs 1000 SEK in salary costs. According to the cost-benefit equation the optimum design is to visit ca 10 sites for both exposed and sheltered shores and to take 2 *Fucus* plants at each.

ANOVA with two orthogonal factors

We have in the previous section seen how nested factors can be included in an experimental design. The nested factors do not represent anything else than another place or another aquarium. The aim is not to study how a certain aquarium may affect the growth of a fish; the aim is only to investigate how much of the variation in growth may be explained by different aquaria. We can think of the aquaria in an experiment as a sample of many possible aquaria. In Part 2 we introduced fixed and random factors and nested factors are typically random factors. Note also that each level of the nested factor *Aquarium* only exists in one level of the *Diet* factor. As we will see below this is an easy way to recognize if a factor is nested. Now we will extend ANOVA to include multiple factors where each level of one factor exists in each level of another factor. Such factors are called orthogonal factors.

Often we have models that include more than one factor and where we are interested to know how the effect of one factor A affects another factor B. The only way to test such complex effects is to design a scientific study where all combinations of factor A and B are included. This is known as a multi-factorial experiment where the factors are orthogonal. The word orthogonal just means that in all levels of factor A all levels of factor B are present. There are two important aspects of analyzing several factors together in a multi-factorial ANOVA:

1. As we will see the statistical power will be much greater than if the factors were analyzed one by one.
2. This is essentially the only statistical technique to test hypotheses about interactions between factors, i.e. how the effect of one factor affects the effect of another factor.

As before we start with an example. A model makes the prediction that induction of a defence compound in a microalgal species occurs when a grazer is present but only when nutrients are scarce. The linear model looks like:

$$X_{ijk} = \mu + G_i + N_j + G*N_{ij} + e_{ijk}$$

The variation in the defence compound is here partitioned due to the two factors *Grazer* and *Nutrient* and also their possible interaction ($G*N_{ij}$). As before we, of course, also have the unexplained variation between individuals, e_{ijk}, not explained by the factors or their interactions. In this model we consider both *Grazer* and *Nutrients* to be fixed factors because *Grazer* contains the 2 levels absence and presence of grazers, and *Nutrients* the 3 levels low, intermediate and high concentration. Would we repeat the experiment we would have selected the same levels (see Part 2 for more information about fixed and random factors). The result of this experiment with 3 replicates of all factor combinations is shown in Fig. 7.

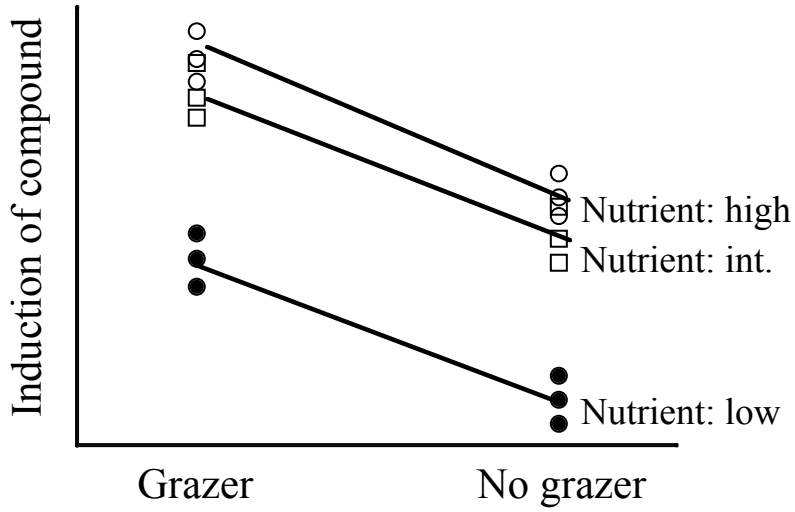

Fig. 7. Result of experiment with two orthogonal factors.

The ANOVA of this multifactorial experiment is seen in Table 8. When we know what variance components that *MS* estimate it is easy to construct the relevant *F*-ratios. In Table 8 both factors and the interaction are all tested against $MS_{residual}$. Remember that the appropriate denominator in an *F*-ratio the *MS* that contains all the variance components except the one tested.

Table 8. The general outline of a 2-factor ANOVA with fixed factors.

Source of Variation	*df*	*MS estimates*
Grazer	a-1	$\sigma^2_e + \frac{nb\sum A_i^2}{a-1}$
Nutrient	b-1	$\sigma^2_e + \frac{na\sum B_i^2}{b-1}$
*Grazer*Nutrient*	(a-1)*(b-1)	$\sigma^2_e + \frac{n\sum AB_{ij}^2}{(a-1)(b-1)}$
Residual	$ab(n$-1)	σ^2_e
Total	abn-1	

Two interesting aspects become apparent with this 2-factor, orthogonal ANOVA:

1. The increase of the statistical power for the main effects (factors *Grazer* and *Nutrient*). The *F*-test for *Grazer* is formed as MS_{Grazer} / $MS_{Residual}$, and $MS_{residual}$ has $ab(n$-1) degrees of freedom, in this example 2*3*2=12. If we had tested *Grazer* in a 1-factor ANOVA $MS_{Residual}$ would have had only $a(n$-1) or 4 *df*. The reason is of course that the test for factor *Grazer* has $b*n$ replicates for each level.
2. We now can test the interaction between the factors *Grazer* and *Nutrient* by forming an *F*-ratio between $MS_{Grazer*Nutrient}$ / $MS_{Residual}$.

The possibility to test interactions between factors is the greatest strength of multifactorial, orthogonal ANOVAs. Interactions are very common in biological systems where complex dynamics and many feedbacks operate. What does a significant interaction imply? In Fig. 7 factors *Grazer* and *Nutrient* both explain a significant amount of the variation in the induction of the defence compound, i.e. we reject H_0 that there is no treatment effect. However, in Fig. 7 there is no significant interaction between *Grazer* and *Nutrient*. In other words, the effect of *Grazer* is the same in all levels of *Nutrient*, and the effect of *Nutrient* is the same in all levels of *Grazer*. The effect of one factor is independent of the other factor. This is further seen as the lines linking the different *Nutrient* levels are parallel.

Figure 8 shows a different result where there is now a significant interaction. Here we clearly see that the effect of our *Grazer* factor is very dependent on the nutrient level; it is only when nutrients are low that the presence of a grazer induces production of the compound. Thus, the reality seems more complicated than in Fig. 7, but we now have a technique to discover it. In Fig. 9 we see yet another example where there is a very

.

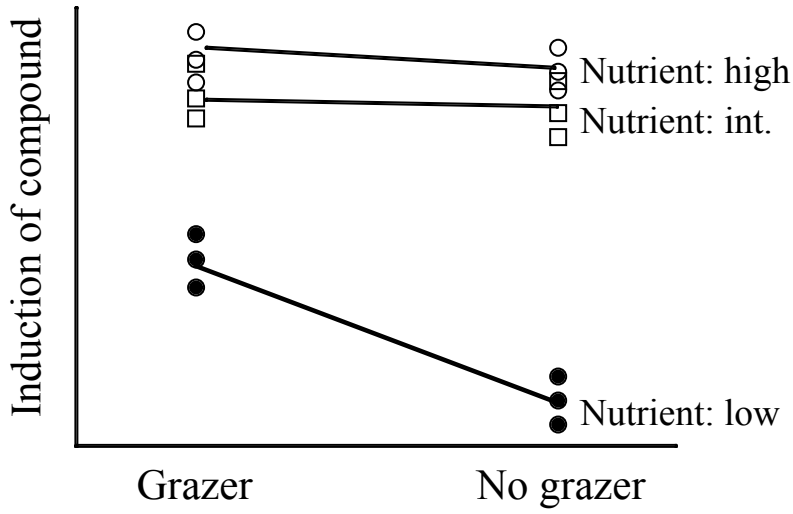

Fig. 8. Result from the 2-factor ANOVA with a significant interaction.

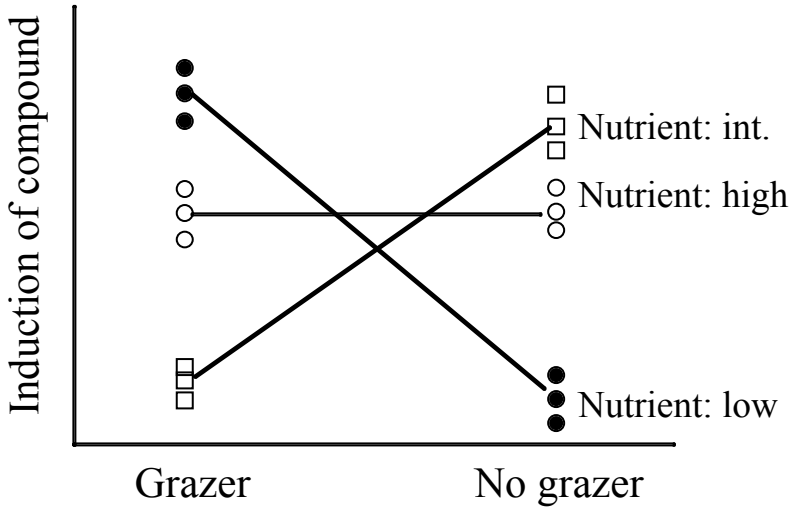

Fig. 9. A 2-factor ANOVA with very strong interaction between the factors.

strong interaction between the 2 factors. In fact, in this example none of the 2 main factors are significant (H_0 is retained). This illustrates that when there is an interaction it is usually not informative to interpret the main factors in isolation. In Fig. 9 this is very clear but even in Fig. 8 the statement that the presence of grazers induces the studied compound is incomplete because this is not a general result since it depends on the nutrient availability. Because of this we always start to test the interaction source of variation in a multi-factorial ANOVA. If the interaction is significant we probably want to know more about how one factor depends on the other. This is usually carried out in a multi-comparison test of the different means of the combinations of the factors. In our example we could test for

a difference between the means of absence and presence of grazers in each nutrient level, and continue with a comparison of the 3 means of nutrient levels when grazers are absent and present, respectively. One example of multi-comparison tests that we have already met is the *a posteriori* SNK-test which can be extended to many factors.

Without going into any detail I want to mention that there is an alternative to *a posteriori* multi-comparison tests of means that is very powerful. Not surprisingly, these tests are called *a priori* tests. The major philosophical difference between *a priori* and *a posteriori* tests is that to make an *a priori* test you must specify a subset of comparisons you want to perform before you carry out the experiment. In the example with the SNK-test above we was potentially interested in all possible differences. In contrast, an *a priori* test is linked to specific hypotheses about the means in the experiment, hypotheses that we have stated before the experiment is actually done. An example in the 2-factor ANOVA above would be that the model we propose to explain the variation of defence compounds predicts that an effect of a grazer would only be detected at low nutrient concentrations. Relevant *a priori* comparisons should show the following patterns if our model is to gain any support:

$\overline{X}_{grazer} > \overline{X}_{no\ grazer}$ in low *Nutrient* and $\overline{X}_{grazer} \leq \overline{X}_{no\ grazer}$ in high *Nutrient*

More than 2 orthogonal factors

A further strength of ANOVA is that it can be extended to analyze very complex experimental designs and sampling programs. It is possible to mix orthogonal with nested factors, and fixed with random factors. An example may be to extend the test of fish growth rate in different diets. We could extend this with one more fixed factor, e.g. *Sex* (female/male) and the random factor *Temperature* with 3 randomly selected levels between 10 and 20°. If the factors *Diet*, *Sex* and *Temperature* are all orthogonal, i.e. all factor combinations are present and that each aquarium represents one such combination we arrive at the design and ANOVA in Table 9. It may look a bit frightening, especially to understand what the *MS* really estimate. We need to be absolutely certain what the different *MS* estimate otherwise we cannot form the appropriate *F*-ratios. This is more complex when we mix fixed and random factors, but as we will see there are rules we can apply in steps to make this fairly easy. A warning may here be justified: many statistical software that analyze ANOVAs with computers do not find the correct *F*-ratios. Thus it is important

Table 9. The outline of an ANOVA with three orthogonal factors and one nested. *Diet* and *Sex* are fixed and *Temperature* and *Aquarium* are random factors. Table continues on next page.

Source of Variation	*df*	*MS estimates*
Diet	a-1	$\sigma^2_e + \sigma^2_{D*T} + \sigma^2_{A(D,S,T)} + k^2_D$
Sex	b-1	$\sigma^2_e + \sigma^2_{S*T} + \sigma^2_{A(D,S,T)} + k^2_S$

Temp.	*c*-1	$\sigma^2_e + \sigma^2_{A(D,S,T)} + \sigma^2_T$
*Diet*Sex*	(*a*-1)(*b*-1)	$\sigma^2_e + \sigma^2_{D*S*T} + \sigma^2_{A(D,S,T)} + k^2_{D*S}$
*Diet*Temp.*	(*a*-1)(*c*-1)	$\sigma^2_e + \sigma^2_{D*T} + \sigma^2_{A(D,S,T)} + \sigma^2_{D*T}$
*Sex*Temp.*	(*b*-1)(*c*-1)	$\sigma^2_e + \sigma^2_{A(D,S,T)} + \sigma^2_{S*T}$
*Diet*Sex*Temp.*	(*a*-1)(*b*-1)(*c*-1)	$\sigma^2_e + \sigma^2_{A(D,S,T)} + \sigma^2_{D*S*T}$
Aquarium (*Diet*, *Sex*, *Temp.*)	*abc*(*d*-1)	$\sigma^2_e + \sigma^2_{A(D,S,T)}$
Residual	*abcd*(*n*-1)	σ^2_e
Total	*abcdn*-1	

that we learn how to find the correct *F*-ratios and this will be our next task.

General method to determine the variance components estimated by MS

We will now walk through a step-wise method to find the variance components that *MS* estimates.

Table 10. Protocol to find the variance components estimated by each *MS*.

Variance component Type of factor	*Residual*	*ABC*	*BC*	*AC*	*AB*	*C* *r*	*B* *f*	*A* *f*
Source of variation								
A	1	-		1	-			1
B	1	-	1		-		1	
C	1	-	-	-		1		
AB	1	1			1			
AC	1	-		1				
BC	1	-	1					
ABC	1	1						
Residual	1							

1. Start to make a protocol as shown in Table 10. Adjust to the number of factors. We will begin with a 3-factor ANOVA where all factors are orthogonal and where A and B are fixed factors and C is a random factor.
2. Label the factors as fixed (with an *f*) or random (with an *r*) in the second row below the variance components.
3. Next, check which factors are nested and under what factors they are nested. Nested factors are indicated in the top row by setting the factor under which it is nested within brackets. If *C* is nested under *A*, *C* is everywhere written as *C*(*A*). Is

C nested under both *A* and *B* this is then indicated as *C*(*AB*). A nested example is shown below.

4. Now, all sources of variation (*MS*) should contain the *Residual* variation. Indicate this with a "1" in the column for *Residual*.
5. All sources of variation (*MS*) contain its own variance component. Also indicate this with a "1" in the appropriate column.
6. Next step is to decide if any other variance components should be included. We start with factor *A*. We begin by making a small mark (e.g. a dash) for all those components that contain factor *A*. Clearly, these are *AB*, *AC* and *ABC*.
7. For each of the selected sources of variation in point 6 above, we cover over the target factor (here A) and look at the other factors. If <u>any</u> of the factors not covered (*B* in *AB, C* in *AC* and *BC* in *ABC*) is fixed this component should <u>not</u> be included. In contrast, if all of the other factors are random this component is included in the MS and you mark this with a "1" in that column. In our example, we should mark *AC*.

   If one of the factors is <u>nested</u> only look at factors outside the brackets, e.g. if *C*(*A*) is nested under *A* and we have identified *BC*(*A*) as a potential component, it should be included if *B* is random but not if *B* is fixed.
8. Now we are finished! Look at all the "1" in the columns and you can write out the variance components that each *MS* estimates. Finally, try to match the appropriate Mean Squares to form *F*-ratios for the different sources of variation. For example the *F*-ratio for testing factor *A* is MS_A / MS_{AC}.

We will take one more example where the third factor *C* is nested under the fixed factor *A*, i.e. *C*(*A*) and where *B* is a random factor. Table 11 shows the result of applying the rules above. Did you get the same result? Since we included a nested factor *C*(*A*) under *A* some of the interactions disappeared. It is not possible to have an interaction between a nested factor and the factor it is nested under. *C*(*A*) after all means that one level (e.g. one aquarium) only receives one level of *A* and this makes it impossible to estimate an interaction. Also note one important thing in Table 11. <u>It is not possible to find an appropriate *F*-ratio for the test of factor *A*</u>. There is no *MS* that lacks the component for *A* but includes all other components present in MS_A. This is often the case when several factors are included in an design and some factors are random. This makes it very important to check <u>before the experiment is carried out</u> that the relevant tests in a given design really exist. If this is not the case we perhaps must re-design the study. This takes some skill and comes with experience.

Table 11. Protocol to find the variance components estimated by each *MS*.

Variance component Type of factor	*Residual*		*BC(A)*	*AB*	*C(A)* *r*	*B* *r*	*A* *f*
Source of variation							
A	1		1	1	1		1
B	1		1	-		1	
C(A)	1		1		1		
AB	1		1	1			
BC(A)	1		1				
Residual	1						

Design of sample programs and experiments

This module of experimental design and statistical analysis aims to integrate the formulation of models and hypotheses with the design of sampling programs and experiments together within a common statistical framework. This process requires that there are logic links between hypothesis, the design of an experiment to test the hypothesis, and the final conclusion. Here follows some important advice how to design and analyze a scientific study

1. Careful planning in advance is essential. This makes it possible to optimize the design and check that it has the ability to answer the relevant questions.
2. Make sure that there is a clear research model predicting one or several hypotheses. Often this involves the identification of possible explanatory factors.
3. Specify the generality of the hypothesis. Should it apply to a specific geographic area, a specific time of the year, a specific range of temperatures etc?
4. Identify the independent replication unit. Is it the aquarium or the fish in the aquarium? Watch out for dependent replicates, especially when the true replication unit is large, or when repeated measurements are collected from each individual.
5. It is important to include appropriate controls, e.g. procedure controls. An example is a transplantation experiment where plants are moved from area A to area B and from B to A, e.g. to test a model that size depends on the local conditions and not on genotype. Here it is important to include the procedural controls where plants are collected at A and then returned to A and the same for B. This controls for any handling effect. Possibly, we would also include an untouched control without any handling. The neglect to include necessary controls is common and may lead to wrong conclusions where a significant effect of a studied treatment (factor) is

actually caused by other aspects of the manipulation. This is called confounding and prevents logic conclusions about factor effects. This is so important that we will take another example. A model states that the reproduction of a macroalgal species is limited because of grazing on reproductive tissue by snails. On randomly selected plots we remove snails and put net cages around the algae to prevent snails from entering the plots. Other plots are used as untouched controls. Our hypothesis is that plants should have more reproductive tissue within the cages where there are no grazing snails. The results of the experiment indeed support this hypothesis. However, something is missing, what? It turns out that a colleague repeats the experiment but she also includes the necessary procedural control, i.e. cages with some openings to allow access from the snails. This control serves to "unconfound" any effect of grazing from the use of cages. And indeed our colleague finds that the increase of reproductive tissue within the cages also occurred in cages where snails could enter. A new model now suggests that wave exposure may limit fecundity and that the cages offered protection from mechanical breakage.

6. Formulate a statistical linear model and identify fixed and random factors. Is any factor nested under another?
7. Write down an ANOVA table and find the variance components that the *MS* estimate. Is it possible to find *F*-ratios for all the tests you are interested in? If not the experiment may have to be re-designed.
8. Draw graphs of how experimental results or observations should appear to support the tested research hypothesis. This may suggest a series of *a priori* comparisons that can be applied for a targeted test of specific hypotheses.
9. Look at the number of degrees of freedom for the tests important to reject or support the research model. Few degrees of freedom, especially in the denominator strongly suggest that the test has low statistical power and that there is a big risk for a type II error. Even better is to perform a power analysis before the experiment to estimate how many degrees of freedom are needed to detect a specified treatment effect. This requires some information about the size of the *MS* in the denominator of the relevant *F*-ratio (see next section about power analysis)

We are all eager to begin the practical parts of any scientific study. However, it cannot be stressed enough that it is extremely important to go through all the points above before setting up experiments or a sampling program. Lack of necessary controls or missing *F*-ratios for essential tests are often revealed at this planning stage.

The formulation of an ANOVA model beforehand also makes it more obvious how the design may change the logic conclusions of the experiment. For example, if a geographic factor *Site* is a random factor any conclusions about this factor will be more general than if

it is considered as fixed, but the price to pay is often reduced statistical power caused by less degrees of freedom. Planning carefully allows for good guesses of statistical power and cost-benefit analyses to check that the desired goal can be reached with available money and time. Often it becomes obvious that the original design is too ambitious and that it will be impossible to detect the treatment effects necessary to test the model. Maybe it is instead possible to test a more specific model that can be tested with greater statistical power to the price of less generality (e.g. fewer species or a limited geographic range).

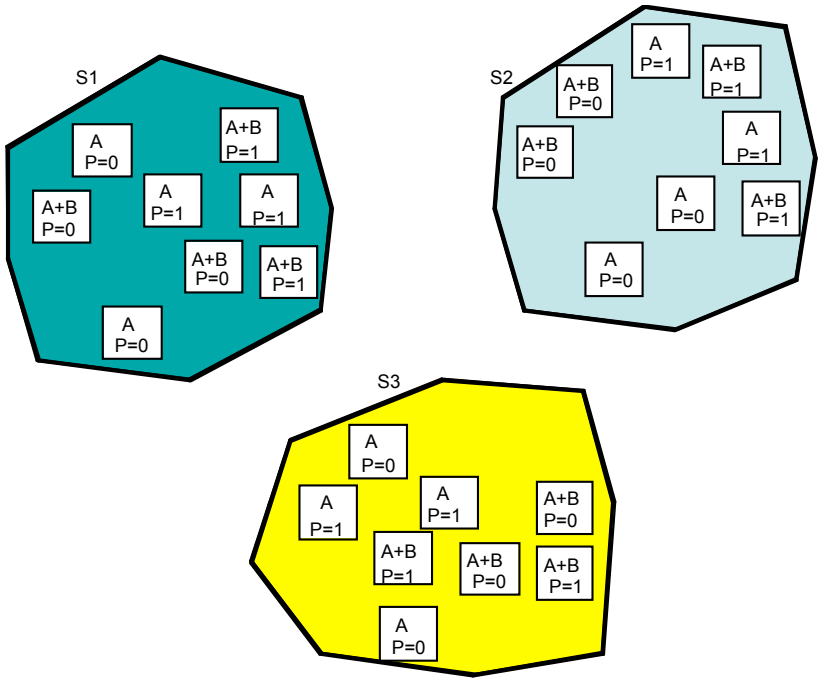

Fig. 10. Schematic outline of field experiment testing effects of predation and competition at three sites. P=1 or 0 indicate presence or absence of predators (manipulated with scarecrows), and A is *L. littorea* only and A+B is *L. littorea* with *L. saxatilis* added.

It takes time to learn how to design logical and efficient experiments. The best way is to practice using real problems. We will next look at three similar field experiments but where the details of the design differs with important implications for the conclusions drawn. The Figs. 10, 11 and 12 shows the outline of field experiments designed to test hypotheses that explain the variation in survival rate of the common periwinkle (*Littorina littorea*). The factors included are in all three cases *Predation* (from birds), *Competition* (from *Littorina saxatilis*) and *Site*. A number of experimental plots are prepared to allow the manipulation of the factors *Predation* and *Competition*. However, the way these plots receive the different treatments differ strongly among the experiments.

In Fig. 10 *Site* is a fixed factor and it represents three islands that are all selected for a Marine Protected Area (MPA) where the knowledge gained from this and other experiments will form an important part of the management of the MPA. The fixed factors *Predation* and *Competition* are both orthogonal to each other and to *Site*. All plots are

completely randomized in each *Site*. There are two independent replicates for each factor combination. The linear model is:

$$X_{ijkl} = \mu + P_i + C_j + S_k + P*C_{ij} + P*S_{ik} + C*S_{jk} + P*C*S_{ijk} + e_{ijkl}$$

By using the rules of finding the variance components estimated by each MS we can write down the ANOVA in Table 12 (try for yourself).

Table 12. ANOVA of the design shown in Fig. 10.

Source of variation	***df***	***MS estimates***
Predation	1	$\sigma^2_e + k^2_P$
Competition	1	$\sigma^2_e + k^2_C$
Site	2	$\sigma^2_e + k^2_S$
*P*C*	1	$\sigma^2_e + k^2_{P*C}$
*P*S*	2	$\sigma^2_e + k^2_{P*S}$
*C*S*	2	$\sigma^2_e + k^2_{C*S}$
*P*C*S*	2	$\sigma^2_e + k^2_{P*C*S}$
Residual	12	σ^2_e
Total	23	

In Fig. 11 the factor *Site* represents something else. Here *Site* is a random factor sampled among many possible sites within the geographic area where we want our conclusions to apply. At every level of *Site* there are two plots receiving one level of the *Predation* factor and one level of the *Competition* factor. Each level of *Site* contains only one level of the other factors and consequently *Site* is nested under both *Predation* and *Competition* (compare the case with aquaria and fish above). The independent experimental unit is each *Site* and the two plots within a site are dependent and can only be used to test if different sites differ. The design in Fig. 11 is common when an increase in generality is desired. Any conclusions about *Predation* and *Competition* can now be logically applied to the whole geographic area from where the different sites were sampled. The linear model is here:

$$X_{ijkl} = \mu + P_i + C_j + S(P, C)_{k(ij)} + P*C_{ij} + e_{ijkl}$$

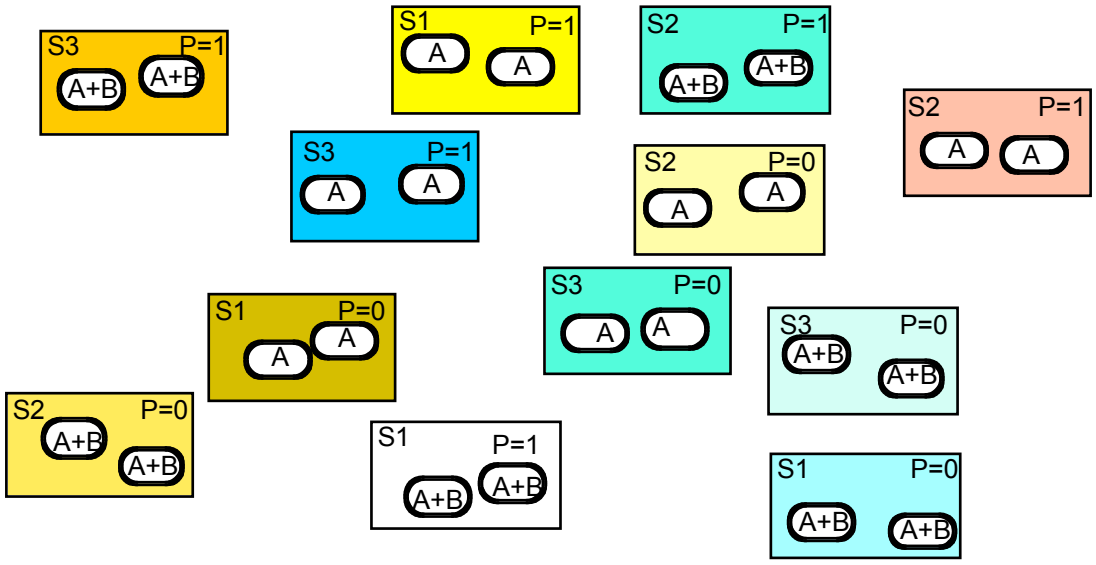

Fig. 11. Schematic outline of field experiment testing effects of predation and competition at 12 sites. Only one combination of *Predation* and *Competition* is present at each site.

Again note that a factor (like *Site*) that is nested under another factor never can be used to estimate their interaction. Thus there are no interaction terms including *Site*. Table 13 shows the ANOVA and the variance components estimated by the *MS*.

Table 13. ANOVA of the design shown in Fig. 11.

Source of variation	***df***	***MS estimates***
Predation	1	$\sigma^2_e + \sigma^2_{S(P,C)} + k^2_P$
Competition	1	$\sigma^2_e + \sigma^2_{S(P,C)} + k^2_C$
*P*C*	1	$\sigma^2_e + \sigma^2_{S(P,C)} + k^2_{P*C}$
Site (*P*, *C*)	8	$\sigma^2_e + \sigma^2_{S(P,C)}$
Residual	12	σ^2_e
Total	23	

Table 13 shows that the price for the increased generality in space by making *Site* a random factor is that the tests of the factors *Predation* and *Competition* now only have 1 over 8 degrees of freedom.

Figure 12 shows an even more intricate design. Also in this design *Site* is a random factor and only represents a sample of different places. As in Fig. 11 there is only one level of *Competition* in each level of *Site*, so *Site* is nested under *Competition*. However, *Site* is orthogonal to *Predation* where both levels are present at each level. This design may be

preferred if there is a lot of work to manipulate plots with only *L. littorea* (A) and both *L. littorea* and *L. saxatilis* (A+B). The linear model that describes this design is:

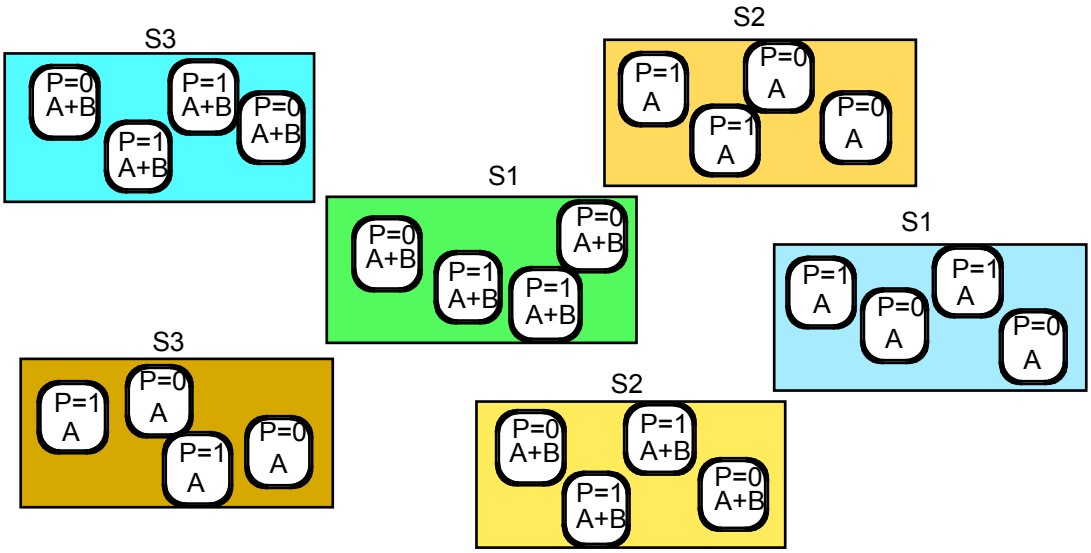

Fig. 12. Schematic outline of field experiment testing effects of predation and competition at 6 sites. Each site contains all levels of Predation but only one level of Competition.

$$X_{ijkl} = \mu + P_i + C_j + S(C)_{k(j)} + P*C_{ij} + P*S(C)_{ik(j)} + e_{ijkl}$$

Table 14 shows the ANOVA and the variance components estimated by *MS*. In Fig. 12 it is sufficient to manipulate the factor *Competition* in six plots. Note, however, that there is a high price to pay because the number of independent experimental units is reduced to only six. It is evident in Table 14 that the test of *Competition* now only involves 1 over 4 degrees of freedom, a weak test indeed. Compare this with the design in Fig. 10 where this test has 1 over 12 *df*. Since *Site* is orthogonal to Predation in Fig. 12 it is now possible to test the interaction between *Predation* and *Site*.

Table 14. ANOVA of the design shown in Fig. 12.

Source of variation	***df***	***MS estimates***
Predation	1	$\sigma^2_e + \sigma^2_{P*S(C)} + k^2_P$
Competition	1	$\sigma^2_e + \sigma^2_{S(C)} + k^2_C$
Site(C)	4	$\sigma^2_e + \sigma^2_{S(C)}$
*P*C*	1	$\sigma^2_e + \sigma^2_{P*S(C)} + k^2_{P*C}$
*P*S(C)*	4	$\sigma^2_e + \sigma^2_{P*S(C)}$
Residual	12	σ^2_e
Total	23	

These three examples show how the design is connected to the analysis, statistical power, and to the logic conclusions that can be drawn.

Analysis of statistical power

In Part 1 we carefully went through the risks of type I and type II errors, and we introduced the concept of statistical power as the probability of rejecting H_0 when H_0 is indeed false. In other words, it indicates the probability to discover an effect, e.g., in an experiment, when there really is an effect. An important part in the testing of research models is to estimate the power of the statistical tests used. Imagine that an ecologically significant effect of predation is at least a 20% reduction in the prey population. An experiment designed to test for such an effect is a waste of time if it is not sufficiently powerful to detect a 20% reduction. Each time we cannot reject H_0 we should consider the statistical power of the test. As already mentioned a power analysis is valuable when we plan an experiment or sampling program. It is then possible in advance to estimate the number of replicates or observations required for the effect size we want to detect. We will end Part 3 with an example of a power analysis.

The objective of a field study is to test the model that sewage treatment plants removing only nitrogen cause higher local biomass of filamentous algae than treatment plants that remove both nitrogen and phosphorous which are expected to be close to the biomass found in non-polluted controls. The requirement (maybe from our client) is that we can detect a change in biomass of 10 g dry weight per m^2 compared to an untreated control area. We have access to a small pilot study (sample size=3) that gave the results shown in Table 15.

Table 15. Biomass of filamentous algae (g dry weight m^{-2})

	Control	Removal of P	Removal of P + N
	10	35	19
	20	18	8
	15	28	25
$\overline{X}_i$	**15**	**27**	**17.3**
s^2	**25**	**73**	**74.3**

We now first begin to calculate the $MS_{Residual}$ that estimates σ^2_e:

$$MS_{residual} = \frac{\sum\sum\left(X_{ij} - \overline{X}_i\right)^2}{a(n-1)} = 57.4$$

The residual *MS* represents the background "noise" of unexplained variation and this determines how strong the treatment effect, the "signal", must be for us to detect it. The power of a test is a function of something called ϕ (phi) which is calculated as:

$$\phi = \sqrt{\frac{n\sum A^2}{aMS_{residual}}}$$ with $a(n\text{-}1)$ degrees of freedom

where n is the number of replicates per treatment, a is the number of treatments (levels) and ΣA^2 is a measure of the treatment effect [$\Sigma A^2=\Sigma(A_i-\overline{X})^2$]. If we want to discover an improvement of 10 g dry weight per m^2 we may imagine a scenario with P-removal 25, NP-removal 15 and control=15 g dry weight per m^2. The summed treatment effect, A^2 is here 67. Statistical power increases with ϕ and the *df*, and we can find the power in statistical tables. It is obvious that power will increase with n and the effect size, and will decrease as $MS_{Residual}$ increases. Table 16 shows the power for different number of replicates. Also note that we have to specify the type I error which here is set to 0.05.

Table 16. Power when a 10 g dry weight m^{-2} must be detected and with α=0.05.

Number of replicates	**Φ**	***df***	***Power***
2	0.76	3	0.13
3	0.93	6	0.24
4	1.08	9	0.35
5	1.20	12	0.46
6	1.32	15	0.56
7	1.42	18	0.65
8	1.52	21	0.72
10	1.62	27	0.83
15	1.71	42	0.95
20	1.87	57	0.99

If we instead are satisfied to detect a difference of 15 g dry weight m^{-2} we get the power in Table 17.

Table 17. Same as in Table 16 but where the effect size is 15 g dry weight m^{-2}.

Number of replicates	**Φ**	***df***	***Power***
2	1.14	3	0.23
3	1.40	6	0.48
4	1.62	9	0.68
5	1.81	12	0.82
6	1.98	15	0.9
7	2.29	18	0.95

We can see from this power analysis that in the final study we need 15 replicates to expect to find a difference of 10 g dry weight m^{-2} with a power of 0.95. If it is sufficient to detect a difference of 15 g dry weight m^{-2} only 7 replicates are needed. In this way the statistical

power can be calculated for most sample programs and experimental designs if the desired treatment effect and the denominator of the F-ratio (here the $MS_{Residual}$) are known.