
"The analysis of variance is more than a technique for statistical analysis. Once it is understood, ANOVA is a tool that can provide an insight into the nature of variation of natural events"

Sokal & Rohlf (1995), BIOMETRY.

3.1. The F distribution

Assume that you are sampling at random from a normally distributed population (or from two different populations with equal variance) by first sampling \( n_1 \) items and calculating their variance \( s_1^2 \) (df: \( n_1 - 1 \)), followed by sampling \( n_2 \) items and calculating their variance \( s_2^2 \) (df: \( n_2 - 1 \)). Now consider the ratio of these two sample variances:

\[
\frac{s_1^2}{s_2^2}
\]

This ratio will be close to 1, because these variances are estimates of the same quantity. The expected distribution of this statistic is called the **F-distribution**. The F-distribution is determined by two values for degrees of freedom, one for each sample variance. The values found within statistical Tables for F (e.g. Table A6) represent \( F_{\alpha(df_1,df_2)} \) where \( \alpha \) is the proportion of the F-distribution to the right of the given F-value and \( df_1, df_2 \) are the degrees of freedom pertaining to the numerator and denominator of the variance ratio, respectively.

![Figure 1](image_url) Three representative F-distributions (note similarity of \( F_{(1,40)} \) to \( \chi^2_1 \)).

For example, a value \( F_{0.025, 9, 9} = 4.03 \) indicates that the ratio \( s_1^2 / s_2^2 \), from samples of ten individuals from normally distributed populations with equal variances, is expected to be larger than 4.03 by chance in only 5% of the experiments (the alternative hypothesis is \( s_1^2 \neq s_2^2 \) so it is a **two tailed test**).
3. 2. Testing the hypothesis of equality of variances

Suppose \( X_1, \ldots, X_m \) are observations drawn from a normal distribution with mean \( \mu_X \) and variance \( \sigma_X^2 \); and \( Y_1, \ldots, Y_n \) are drawn from a normal distribution with mean \( \mu_Y \) and variance \( \sigma_Y^2 \). In theory, the F statistic can be used as a test for the hypothesis \( H_0: \sigma_X^2 = \sigma_Y^2 \) vs. the hypothesis \( H_1: \sigma_X^2 \neq \sigma_Y^2 \). \( H_0 \) is rejected at the \( \alpha \) level of significance if the ratio \( s_X^2/s_Y^2 \) is either \( \geq F_{\alpha/2, dfX-1, dfY-1} \) or \( \leq F_{1-\alpha/2, dfX-1, dfY-1} \). In practice, this test is rarely used because it is very sensitive to departures from normality. This can be calculated using SAS PROC TTEST.

3. 3. Testing the hypothesis of equality of two means

The ratio between two estimates of \( \sigma^2 \) can also be used to test differences between means; that is, it can be used to test \( H_0: \mu_1 - \mu_2 = 0 \) versus \( H_1: \mu_1 - \mu_2 \neq 0 \). In particular:

\[
F = \frac{\text{estimate of } \sigma^2 \text{ from sample means}}{\text{estimate of } \sigma^2 \text{ from individuals}}
\]

The denominator is an estimate of \( \sigma^2 \) provided by the individuals within each sample. That is, it is a weighted average of the sample variances.

The numerator is an estimate of \( \sigma^2 \) provided by the means among samples. The variance of a population of sample means is \( \sigma^2/n \), where \( \sigma^2 \) is the variance of individuals in a parent population and all samples are of size \( n \). This implies that means may be used to estimate \( \sigma^2 \) by multiplying the variance of sample means \( \sigma^2/n \) by \( n \).

\[
F = \frac{s_{\text{among}}^2}{s_{\text{within}}^2} = \frac{n s_T^2}{s^2}
\]

When the two populations have different means (but the same variance), the estimate of \( \sigma^2 \) based on sample means will include a contribution attributable to the difference among population means as well as any random difference (i.e. within-population variance). Thus, in general, if the means differ, the sample means are expected to be more variable than predicted by chance alone.

Example: We will explain the test using a data set of Little and Hills (p. 31). Yields (100 lb./acre) of wheat varieties 1 and 2 from plots to which the varieties were randomly assigned:

<table>
<thead>
<tr>
<th>Varieties</th>
<th>Replications</th>
<th>( Y_i )</th>
<th>( \bar{Y}_i )</th>
<th>( s_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19 14 15 17 20</td>
<td>85</td>
<td>( \bar{Y}_1 = 17 )</td>
<td>6.5</td>
</tr>
<tr>
<td>2</td>
<td>23 19 19 21 18</td>
<td>100</td>
<td>( \bar{Y}_2 = 20 )</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>( Y_\cdot = 185 )</td>
<td>( \bar{Y}_\cdot = 18.5 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this experiment, there are two treatment levels \( t = 2 \) and five replications \( r = 5 \) (the symbol \( t \) stands for "treatments" and \( r \) stands for "replications").

Each observation in the experiment has a unique "address" given by \( Y_{ij} \), where \( i \) is the index for treatment \( (i = 1,2) \) and \( j \) is the index for replication \( (j = 1,2,3,4,5) \). Thus \( Y_{24} = 19 \).

The dot notation is a shorthand alternative to using \( \sum \). Summation is for all values of the subscript occupied by the dot. Thus \( Y_{1.} = 19 + 14 + 15 + 17 + 20 \) and \( Y_{.2} = 14 + 19 \).

We begin by assuming that the two populations have the same (unknown) variance \( \sigma^2 \) and then test \( H_0: \mu_1 = \mu_2 \). We do this by obtaining two estimates for \( \sigma^2 \) and comparing them.

First, we can compute the average variance of individuals within samples, also known as the experimental error. To determine the experimental error, we compute the variance of each sample \( (s^2_1 \text{ and } s^2_2) \), assume they both estimate a common variance, and then estimate that common variance by pooling the two estimates:

\[
\frac{\sum_j (Y_{1j} - \bar{Y}_1)^2}{r_1 - 1}, \quad \frac{\sum_j (Y_{2j} - \bar{Y}_2)^2}{r_2 - 1}
\]

\[
s^2_{\text{pooled}} = \frac{(r_1 - 1)s^2_1 + (r_2 - 1)s^2_2}{(r_1 - 1) + (r_2 - 1)} = \frac{4*6.5 + 4* 4.0}{(4 + 4)} = 5.25 \equiv s^2_{\text{within}}
\]

In this case, since \( r_1 = r_2 \), the pooled variance is simply the average of the two sample variances. Since pooling \( s^2_1 \) and \( s^2_2 \) gives an estimate of \( \sigma^2 \) based on the variability within samples, let's designate it \( s^2_w \) (subscript \( w = \text{within} \)).

The second estimate of \( \sigma^2 \) is based on the variation between or among samples. Assuming, by the null hypothesis, that these two samples are random samples drawn from the same population and that, therefore, \( \bar{Y}_1 \) and \( \bar{Y}_2 \) both estimate the same population mean, we can estimate the variance of means of that population by \( s^2_y \). Recall from Topic 1 that the mean \( \bar{Y} \) of a set of \( r \) random variables drawn from a normal distribution with mean \( \mu \) and variance \( \sigma^2 \) is itself a normally distributed random variable with mean \( \mu \) and variance \( \sigma^2/r \).

The formula for \( s^2_y \) is

\[
s^2_y = \frac{\sum_{i=1}^t (\bar{Y}_i - \bar{Y}).)^2}{r - 1} = \frac{[(17 - 18.5)^2 + (20 - 18.5)^2]}{(2-1)} = 4.5
\]

and, from the central limit theorem, \( n \) times this quantity provides an estimate for \( \sigma^2 \) (\( r \) is the number of variates on which each sample mean is based).
Therefore, the between samples estimate is:

\[ r s_t^2 = 5 \times 4.5 = 22.5 \equiv s_{between} \]

These two variances are used in the F test as follows. If the null hypothesis is not true, we would expect the variance between samples to be much larger than the variance within samples ("much larger" means larger than one would expect by chance alone). Therefore, we look at the ratio of these variances and ask whether this ratio is significantly greater than 1. It turns out that under our assumptions (normality, equal variance, etc.), this ratio is distributed according to an \( F_{(t-1, t(r-1))} \) distribution. That is, we define:

\[ F = \frac{s_b^2}{s_w^2} \]

and test whether this statistic is significantly greater than 1. The F statistic is a measure of how many times larger the variability between the samples is compared to the variability within samples. In this example, \( F = 22.5/5.25 = 4.29 \). The numerator \( s_b^2 \) is based on 1 df, since there are only two sample means. The denominator, \( s_w^2 \), is based on pooling the df within each sample so \( df_{den} = t(n-1) = 2(4) = 8 \). For these df, we would expect an F value of 4.29 or larger just by chance about 7% of the time. From Table A.6, \( F_{0.05, 1, 8} = 5.32 \). Since 4.29 < 5.32, we fail to reject \( H_0 \) at the 0.05 significance level.

### 3.3.1 Relationship between F and t

In the case of only two treatments, the square-root of the F statistic is distributed according to a t distribution:

\[ F_{1-\alpha, df=1, t(r-1)} = t^2 \quad \text{meaning} \quad t = \sqrt{\frac{s_b^2}{s_w^2}} \]

In the example above, with 5 reps per treatment:

\[ F_{(1,8), 1-\alpha} = (t_5, 1-\alpha/2)^2 \]

\[ 5.32 = 2.306^2 \]

The total degrees of freedom for the t statistic is \( t(r - 1) \) since there are rt total observations and they must satisfy t constraint equations, one for each treatment mean. Therefore, we reject the null hypothesis at the \( \alpha \) significance level if \( t > t_{\alpha/2, t(r-1)} \).
Here are the computations for our data set:

\[ t = \frac{s^2_b}{s^2_w} = \frac{22.5}{5.25} = 2.07 \]

Since \( 2.07 < t_{0.025, 8} = 2.306 \), we fail to reject \( H_0 \) at the 0.05 significance level.

### 3.4. The linear additive model

#### 3.4.1. One population: In statistics, a common model describing the makeup of an observation states that it consists of a mean plus an error. This is a linear additive model. A minimum assumption is that the errors are random, making the model probabilistic rather than deterministic.

The simplest linear additive model is this one:

\[ Y_i = \mu + \varepsilon_i \]

It is applicable to the problem of estimating or making inferences about population means and variances. This model attempts to explain an observation \( Y_i \) as a mean \( \mu \) plus a random element of variation \( \varepsilon_i \). The \( \varepsilon_i \)'s are assumed to be from a population of uncorrelated \( \varepsilon \)'s with mean zero. Independence among \( \varepsilon \)'s is assured by random sampling.

#### 3.4.2. Two populations: Now consider this model:

\[ Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \]

It is more general than the previous model because it permits us to describe two populations simultaneously. For samples from two populations with possibly different means but a common variance, any given reading is composed of the grand mean \( \mu \) of the population, a component \( \tau_i \) for the population involved (i.e. \( \mu + \tau_1 = \mu_1 \) and \( \mu + \tau_2 = \mu_2 \)), and a random deviation \( \varepsilon_{ij} \). The subindex \( i \) (= 1, 2) indicates the treatment number and the subindex \( j \) (= 1, ..., \( r \)) indicates the number of observations from each population (replications).

\( \tau_i \), the treatment effects, are measured as deviations from the overall mean \( [\mu = (\mu_1 + \mu_2) / 2] \) such that \( \tau_1 + \tau_2 = 0 \) or \( -\tau_1 = \tau_2 \). This does not affect the difference between means, \( 2\tau \). If \( r_1 \neq r_2 \) we may set \( r_1\tau_1 + r_2\tau_2 = 0 \).

The \( \varepsilon \)'s are assumed to be from a single population with normal distribution, mean \( \mu = 0 \), and variance \( s^2 \).
Another way to express this model, using the dot notation from before, is:

\[ Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.}) \]

### 3. 4. 3. More than two populations. One-way classification ANOVA

As with the 2 sample t-test, the linear model is:

\[ Y_{ij} = \mu + \tau_i + \epsilon_{ij} \]

where now \( i = 1, \ldots, t \) and \( j = 1, \ldots, r \). Again, the \( \epsilon_{ij} \) are assumed to be drawn from a normal distribution with mean 0 and variance \( \sigma^2 \). Two different kinds of assumptions can be made about the \( \tau_i \)'s that will differentiate the Model I ANOVA from the Model II ANOVA.

**The Model I ANOVA or fixed model:** In this model, the \( \tau_i \)'s are fixed and

\[ \sum \tau_i = 0 \]

The constraint \( \sum \tau_i = 0 \) is a consequence of defining treatment effects as deviations from an overall mean. The null hypothesis is then stated as \( H_0: \tau_1 = \ldots = \tau_i = 0 \) and the alternative as \( H_1: \) at least one \( \tau_i \neq 0 \). What a Model I ANOVA tests is the differential effects of treatments that are fixed and determined by the experimenter. The word "fixed" refers to the fact that each treatment is assumed to always have the same effect \( \tau_i \). Moreover, the set of \( \tau_i \)'s are assumed to constitute a finite population and are specific parameters of interest, along with \( \sigma^2 \). In the case of a false \( H_0 \) (i.e. some \( \tau_i \neq 0 \)), there will be an additional component of variation due to treatment effects equal to:

\[ r \sum \frac{\tau_i^2}{t-1} \]

Since the \( \tau_i \) are measured as deviations from a mean, this quantity is analogous to a variance but cannot be called such since it is not based on a random variable but rather on deliberately chosen treatments.

**The Model II ANOVA or random model:** In this model, the additive effects for each group (\( \tau_i \)'s) are not fixed treatments but are random effects. In this case, we have not deliberately planned or fixed the treatment for any group, and the effects on each group are random and only partly under our control. The \( \tau_i \) themselves are a random sample from a population of \( \tau_i \)'s for which the mean is zero and the variance is \( \sigma^2 \). When the null hypothesis is false, there will be an additional component of variance equal to \( r\sigma^2 \). Since the effects are random, it is futile to estimate the magnitude of these random effects for any one group or the differences from group to group; but we can estimate their variance, the added variance component among groups: \( r\sigma^2 \). We test for its presence and estimate its magnitude, as well as its percentage contribution to the variation. In
the fixed model, we draw inferences about particular treatments; in the random model, we draw an inference about the population of treatments. The null hypothesis in this latter case is stated as \( H_0: \sigma^2_t = 0 \) versus \( H_1: \sigma^2_t \neq 0 \).

An important point is that the basic setup of data, as well as the computation and significance test, in most cases is the same for both models. It is the purpose which differs between the two models, as do some of the supplementary tests and computations following the initial significance test. For now, we will deal only with the fixed model.

**Assumptions of the model**  [ST&D 174]

1. Treatment and environmental effects are additive
2. Experimental errors are random, possess a common variance, and are independently and normally distributed about zero mean

**Effects are additive**
This means that all effects in the model (treatment effects, random error) cause deviations from the overall mean in an additive manner (rather than, for example, multiplicative).

**Error terms are independently and normally distributed**
This means there is no correlation between experimental groupings of observations (e.g. by treatment level) and the sizes of the error terms. This could be violated if, for example, treatments are not assigned randomly.

This assumption essentially means that the means and variances of treatments share no correlation. For example, suppose yield is measured and the treatments cause yield to range from 1 g/plant up to 10 g/plant. A range of ± 1 gm would be much more "significant" at the low end than the high end but could not be considered any differently by this model.

**Variances are homogeneous**
The means the variances of the different treatment groups are the same.
3.5. ANOVA: Single factor designs

3.5.1. The Completely Random Design CRD

In single factor experiments, a single treatment (i.e. factor) is varied to form the different treatment levels. The experiment discussed below is taken from page 141 of ST&D. The experiment involves inoculating five different cultures of one legume, clover, with strains of nitrogen-fixing bacteria from another legume, alfalfa. As a sort of control, a sixth trial was run in which a composite of five clover cultures was inoculated. There are 6 treatments (t = 6) and each treatment is given 5 replications (r = 5).

<table>
<thead>
<tr>
<th></th>
<th>3DOK1</th>
<th>3DOK5</th>
<th>3DOK4</th>
<th>3DOK7</th>
<th>3DOK13</th>
<th>composite</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{ij}</td>
<td>144.1</td>
<td>119.9</td>
<td>73.2</td>
<td>99.6</td>
<td>66.3</td>
<td>93.5</td>
<td>596.6</td>
</tr>
<tr>
<td>$\bar{Y}_i$</td>
<td>4287.53</td>
<td>2932.27</td>
<td>1139.42</td>
<td>1989.14</td>
<td>887.29</td>
<td>1758.71</td>
<td>12994.36</td>
</tr>
<tr>
<td>$\bar{Y}_i/r$</td>
<td>4152.96</td>
<td>2875.2</td>
<td>1071.65</td>
<td>1984.03</td>
<td>879.14</td>
<td>1748.45</td>
<td>12711.43</td>
</tr>
<tr>
<td>$\sum (Y_{ij} - \bar{Y}_i)^2$</td>
<td>134.57</td>
<td>57.07</td>
<td>67.77</td>
<td>5.11</td>
<td>8.15</td>
<td>10.26</td>
<td>282.93</td>
</tr>
<tr>
<td>$\bar{Y}_i$ - mean</td>
<td>28.8</td>
<td>24.0</td>
<td>14.6</td>
<td>19.9</td>
<td>13.3</td>
<td>18.7</td>
<td>19.88</td>
</tr>
<tr>
<td>$\sigma^2_{n-1}$</td>
<td>33.64</td>
<td>14.27</td>
<td>16.94</td>
<td>1.28</td>
<td>2.04</td>
<td>2.56</td>
<td></td>
</tr>
</tbody>
</table>

Inoculation of clover with Rhizobium strains [ST&D Table 7.1]

The completely randomized design (CRD) is the basic ANOVA design. It is used when there are t different treatment levels of a single factor (in this case, Rhizobium strain). These treatments are applied to t independent random samples of size r. Let the total sample size for the experiment be designated as n = rt. Let $Y_{ij}$ denote the jth measurement (replication) recorded from the ith treatment. WARNING: Some texts interchange the i and the j (i.e. the rows and columns of the table), so be careful.

We wish to test the hypothesis $H_0$: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_t$ against $H_1$: not all the $\mu_i$'s are equal.

This is a straightforward extension of the two-sample t test of topic 3.3 since there was nothing special about the value t = 2. Recall that the test statistic was:

$$ F = \frac{s_b^2}{s_w^2} $$

In our new dot notation, we can write:

$$ s_w^2 = \frac{\sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2}{t(r-1)} = \frac{SSE}{t(r-1)} $$

where $SSE \equiv \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2$
Here SSE is the **sum of squares for error**. Also:

\[
\sum_{i=1}^{t} (Y_i - \bar{Y})^2
\]

where \( \text{SST} = \sum_{i=1}^{t} (Y_i - \bar{Y})^2 \)

Here SST is the **sum of squares for treatment** (SAS refers to this as the Model SS).

Since the variance among treatment means estimates \( \sigma^2 / r \), the \( r \) in the definition formula for SST is required so that the **mean square for treatment** (MST) will be an estimate of \( \sigma^2 \) rather than \( \sigma^2 / r \). This is equivalent to the step we took in example 3.3 above when we multiplied by \( r \) in order to estimate the variance between samples (\( s_b^2 = r s_y^2 \)).

Using our new notation, we can write:

\[
F = \frac{\text{SST} / (t-1)}{\text{SSE} / (n-t)} = \frac{\text{SST} / (t-1)}{\text{SSE} / (n-1)}
\]

We can then define:

**The mean square for error**: \( \text{MSE} = \text{SSE} / (n-t) \). This is the average dispersion of the observations around their respective group means. It is an estimate of a common \( \sigma^2 \), the experimental error (i.e. the variation among observations treated alike). MSE is a valid estimate of the common \( \sigma^2 \) if the assumption of equal variances among treatments is true.

**The mean square for treatment**: \( \text{MST} = \text{SST} / (t-1) \). (MS Model in SAS) This is an independent estimate of \( \sigma^2 \), when the null hypothesis is true (\( H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t \)). If there are differences among treatment means, there will be an additional source of variation in the experiment due to treatment effects equal to \( r \sum \tau_i^2 / (t-1) \) (Model I) or \( r \sigma_i^2 \) (Model II) (see topic 3.4.3 and ST&D 155).

\[
F = \frac{\text{MST}}{\text{MSE}}
\]

The \( F \) value is obtained by dividing the treatment mean square by the error mean square. We expect to find \( F \) approximately equal to 1. In fact, however, the expected ratio is:

\[
\frac{\text{MST}}{\text{MSE}} = \frac{\sigma^2 + r \sum \tau_i^2 / (t-1)}{\sigma^2}
\]

As is clear from this formula, the \( F \)-test is sensitive to the presence of the added component of variation due to treatment effects. In other words, ANOVA permits us to test whether there are
any nonzero treatment effects. That is, to test whether a group of means can be considered random samples from the same population or whether we have sufficient evidence to conclude that the treatments that have affected each group separately have resulted in shifting these means sufficiently so that they can no longer be considered samples from the same population.

Recall that the degrees of freedom is the number of independent, unconstrained quantities underlying a statistic.

Underlying SST are t quantities \((\bar{Y}_i - \bar{Y}_.)\) which have one constraint (they must sum to 0); so \(df_{trt} = t-1\). Underlying SSE are n quantities \(Y_{ij}\), which have t constraints for the t sample means; so \(df_e = n-t\).

Consider the following equation:

If you take the time to deal with the messy algebra, you will find that this equality is true. The reason this is relevant is because this is just our dot notation for:

\[
TSS = SST + SSE
\]

where TSS is the total sum of squares of the experiment. In other words, sums of squares are perfectly additive. If you were to fully expand the quantity on the left-hand side of the equation, you find lots of cross product terms of the form \(2(\bar{Y}_{ij} \bar{Y}_.)\). It turns out, upon simplification, that all of those cross product terms cancel each other out (note that none appear on the right-hand side of the equation). Quantities that satisfy this criterion are said to be orthogonal. Another way of saying this is that we can decompose the total SS into a portion due to variation among groups and another, completely independent portion due to variation within groups. The degrees of freedom are also additive (i.e. \(df_{Tot} = df_{Trt} + df_e\).
The dot notation above provides the "definitional" forms of these quantities (TSS, SST, and SSE). But each also has a friendlier "calculational" form, for when you compute them by hand. The two expressions are mathematically equivalent. The calculational forms make use of a quantity called the "correction factor":

\[ C = \frac{1}{n} (\overline{Y})^2 = \frac{1}{n} \left( \sum_{ij} Y_{ij} \right)^2 \]

This term is just the squared sum of all observations in the experiment, divided by their total number. Once you calculate C, you can tackle the total SS (TSS):

\[ TSS = \left( \sum_{i=1}^{t} \sum_{j=1}^{r} Y_{ij}^2 \right) - C \]

The total SS is the sum of squares that includes all sources of variation. In the dot notation, you see that it is the sum of the squares of the deviation of each observation from the overall mean.

Next, you can tackle the treatment sum of squares (SST):

\[ SST = \frac{1}{r} \left( \sum_{i=1}^{t} Y_{i.}^2 \right) - C \]

The SST is the sum of squares attributable to the variable of classification. This is the SS due to differences among treatment groups and is referred to as the within-or-among groups SS.

Finally, the error sum of squares (SSE):

\[ SSE = TSS - SST \]

The SSE is that part of the total sum of squares that cannot be explained by difference among groups. It is the sum of squares among individuals treated alike. It is referred to as the within groups SS, residual SS, or error SS.
An ANOVA table provides a systematic presentation of everything we've covered until now. The first column of the ANOVA table specifies the components of the linear model. In a single factor CRD, remember, the linear model is just:

\[ Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \]

By this model, we have two named sources of variation: Treatments and Error. The next column indicates the df associated with each of these components. Next is a column with the SS associated with each, followed by a column with the mean squares associated with each. Mean squares, mathematically, are essentially variances; and they are found by dividing SS by their respective df. Finally, the last column in an ANOVA table present the F statistic, which is a ratio of mean squares (i.e. a ratio of variances).

An ANOVA table (including an additional column of the SS definitional forms):

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Definition</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>t - 1</td>
<td>( \sum_i (\bar{Y}_i - \bar{Y})^2 )</td>
<td>SST</td>
<td>SST/(t-1)</td>
<td>MST/MSE</td>
</tr>
<tr>
<td>Error</td>
<td>t(r-1) = n - t</td>
<td>( \sum_{i,j} (Y_{ij} - \bar{Y}_i)^2 )</td>
<td>TSS - SST</td>
<td>SSE/(n-t)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>n - 1</td>
<td>( \sum_{i,j} (Y_{ij} - \bar{Y})^2 )</td>
<td>TSS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ANOVA table for our Rhizobium experiment would look like this:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>5</td>
<td>847.05</td>
<td>169.41</td>
<td>14.37**</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>282.93</td>
<td>11.79</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>1129.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notice that the mean square error (MSE = 11.79) is just the pooled variance or the average of variances within each treatment (i.e. MSE = \( \Sigma s_i^2 / t \); where \( s_i^2 \) is the variance estimated from the ith treatment). The F value of 14 indicates that the variation among treatments is over 14 times larger than the mean variation within treatments. This value far exceeds the critical F value for such an experimental design at \( \alpha = 0.05 (F_{\text{crit}} = F_{(5,24),0.05} = 2.62) \), so we reject \( H_0 \). At least one of the treatments has a nonzero effect on the response variable, at the specified significance level.
3.5.1.2. Assumptions associated with ANOVA

The assumptions associated with ANOVA can be expressed in terms of the following statistical model:

\[ Y_{ij} = \mu + \tau_i + \epsilon_{ij}. \]

First, \( \epsilon_{ij} \) (the residuals) are assumed to be **normally distributed** with mean 0 and possess a **common variance** \( \sigma^2 \), independent of treatment level \( i \) and sample number \( j \).

3.5.1.2.1. Normal distribution

Recall from the first lecture that the Shapiro-Wilk test statistic W (ST&D 567; produced by SAS via Proc UNIVARIATE NORMAL) provides a powerful test for normality for small to medium samples (\( n < 2000 \)).

For large populations (\( n > 2000 \)), SAS recommends the use of the Kolmogorov-Smirnov statistics (ST&D 571; produced by SAS via Proc UNIVARIATE NORMAL or via Analyst).

3.5.1.2.2. Homogeneity of treatment variances

Tests for homogeneity of variance (i.e. homoscedasticity) attempt to determine if the variance is the same within each of the groups defined by the independent variable. Bartlett's test (ST&D 481) can be very inaccurate if the underlying distribution is even slightly nonnormal, and it is not recommended for routine use. Levene's test is much more robust to deviations from normality.

Levene's test is an ANOVA of the squares of the residuals of each observation from its treatment mean. If Levene's test leads you reject the null hypothesis (\( H_0 \): the mean squared residual is the same for all treatments; i.e. the within-treatment variance is the same across all treatment groups; i.e. variances are homogeneous) the Analyst Application provides the option of performing a Welch's variance-weighted ANOVA (Biometrika 1951 v38, 330) instead of the usual ANOVA to test for differences between group means. This alternative to the usual analysis of variance is more robust if variances are not equal.

To perform Levene's test in SAS, you use the option HOVTEST (for Homogeneity of variance test) within the means statement in the PROC GLM procedure:

```
Proc GLM;
  Class Treatment;
  Model Response = Treatment;
  Means Treatment / HOVTEST = Levene;
```
3.5.1.3. Experimental Procedure: Randomization

Here is how the clover plots might look if this experiment were conducted in the field:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

The experimental procedure would be: First, randomly (e.g. from a random number table such as ST&D 606, or using PROC PLAN in SAS, etc.) select the plot numbers to be assigned to the six treatments (A,B,C,D,E,F). Example: On p. 607, starting from Row 02, columns 88-89 (a random starting point), move downward. Take for treatment A the first 5 random numbers under 30, and so forth: Treatment A: 05, 19, 13, 20, 6; Treatment B: 14, 26, 1, 8, 4; etc.

3.5.1.4. Power and sample size

3.5.1.4.1. Power

The power of a test is the probability of detecting a nonzero treatment effect. To calculate the power of the F test in an ANOVA using Pearson and Hartley's power function charts (1953, Biometrika 38:112-130), it is necessary first to calculate a critical value \( \phi \). This critical value depends on the number of treatments (t), the number of replications (r), the magnitude of the treatment effects that the investigator wishes to detect (d), an estimate of the population variance (\( \sigma^2 = \text{MSE} \)), and the probability of rejecting a true null hypothesis (\( \alpha \)).

In a CRD, \( y_{ij} = \mu + \tau_i + \varepsilon_{ij} \),

where \( i = 1,2,\ldots,t; j = 1,2,\ldots,r; \mu \) is the overall mean; and \( \tau_i \) is the treatment effect (\( \tau_i = \mu_i - \mu \)). To calculate the power, you first need to calculate \( \phi \), a standardized measure (in \( \sigma \) units) of the expected differences among means which can be used to determine sample size from the power charts. Its general form:

\[
\phi = \sqrt{\frac{r}{\text{MSE}}} \sum_{i} \frac{\tau_i^2}{t}
\]

This general formula can be simplified if we assume all \( \tau_i \) are zero except the two extreme treatment effects (let's call them \( \tau_K \) and \( \tau_L \), so that \( d = |\mu_K - \mu_L| \)). You can think of \( d \) as the difference between the extreme treatment means.
Taking $\mu$ to be in the middle of $\mu_K$ and $\mu_L$, $\tau_i = d/2$:

$$\sum \frac{\tau_i^2}{t} = \frac{(d/2)^2 + (d/2)^2}{t} = \frac{d^2/4 + d^2/4}{t} = \frac{d^2/2}{t} = \frac{d^2}{2t}$$

And the $\phi$ formula simplifies:

$$\phi = \sqrt{\frac{d^2 \cdot r}{2t \cdot MSE}}$$

With this value, you enter the chart for $v_1 = df_1 = df_{\text{numerator}} = df_{\text{treatment}} = t-1$ and choose the x-axis scale for the appropriate $\alpha$ (0.05 or 0.01). The interception of the calculated $\phi$ with the curve for $v_2 = df_2 = df_{\text{denominator}} = df_{\text{error}} = t(n-1)$ gives the power of the test (the y-axis on both sides of the chart).

**Example:** Suppose an experiment has $t = 6$ treatments with $r = 2$ replications each. The difference between the extreme means is 10 units, MSE= 5.46, and the required $\alpha = 5\%$. To calculate the power using the simplified formula:

$$\phi = \sqrt{\frac{d^2 \cdot r}{2t \cdot MSE}} = \sqrt{\frac{10^2 \cdot 2}{2(6) \cdot 5.46}} = 1.75$$

Use Chart $v_1 = t-1 = 5$ and the set of curves to the left ($\alpha = 5\%$). Select curve $v_2 = t(r-1) = 6$. The height of this curve corresponding to the abscissa of $\phi = 1.75$ is the power of the test. In this case, the power is slightly greater than 0.55. As a rule of thumb, experiments should be designed with a power of at least 80% (i.e. $\beta \leq 0.20$).

To calculate the power using *Analyst:* Statistics $\rightarrow$ ANOVA $\rightarrow$ One-Way ANOVA $\rightarrow$ Tests $\rightarrow$ Power analysis. Or *Analyst* $\rightarrow$ Sample Size $\rightarrow$ One-Way ANOVA $\rightarrow$ Complete the number of treatments, the corrected sum of squares CSS (= SST = between SS = among groups SS = treatment SS), and the standard deviation, which is the square root of the mean squared error (MSE). You must also specify the significance level of the test; the default is 0.05.

**3. 5. 1. 4. 2. Sample size**

To calculate the number of replications for a given $\alpha$ and desired power: A) Specify the constants, B) Start with an arbitrary $r$ to compute $\phi$, C) Use the appropriate Pearson and Hartley chart to find the power; and D) Iterate the process until a minimum $r$ value is found which satisfies the required power for a given $\alpha$ level.
Example: Suppose that 6 treatments will be involved in a study and the anticipated difference between the extreme means is 15 units. What is the required sample size so that this difference will be detected at \( \alpha = 1\% \) and power = 90\%, knowing that \( \sigma^2 = 12 \)? (note, \( t = 6, \ \alpha = 0.01, \beta = 0.10, \ d = 15, \ \text{and MSE} = 12 \)).

<table>
<thead>
<tr>
<th>r</th>
<th>df</th>
<th>( \phi )</th>
<th>(1-( \beta )) for ( \alpha=1% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6(2-1)= 6</td>
<td>1.77</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>6(3-1)= 12</td>
<td>2.17</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>6(4-1)= 18</td>
<td>2.50</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Thus 4 replications are required for each treatment to satisfy the required conditions.

3.5.2. Subsampling: the nested design

It may happen that the experimenter wishes to make several observations within each experimental unit, the unit to which the treatment is applied. Such observations are called subsamples. The classical example of this is given in Steel and Torrie: sampling individual plants within pots where the pots are the experimental units randomly assigned to treatments. Other examples would be individual trees within an orchard plot (where the treatment is assigned to the plot), individual sheep within a herd (where the treatment is assigned to the herd), etc. We call the analysis of this kind of data a nested analysis of variance. Nested ANOVAs are not limited to two hierarchical levels (e.g. pots, and then plants within pots). We can divide the subgroups into sub-subgroups, and even further, as long as the sampling units within each level are chosen randomly (e.g. pots, then plants within pots, then flowers within plants, then anthers within flowers, etc. etc.).

The essential objective of a nested ANOVAs is to dissect the MSE of a system into its components, thereby ascertaining the sources and magnitudes of error in an experiment or process. One example of this objective would be to discover and characterize sources of variation in systematic studies of natural populations.

3.5.2.1. Linear model for subsampling

Before we perform a nested ANOVA, let us examine the linear model upon which it is based:

\[ Y_{ijk} = \mu + \tau_i + \varepsilon_{j(i)} + \delta_{k(ij)} \]

The interpretations of \( \mu, \ \tau, \ \text{and} \ \varepsilon \) are as before. But now two random elements are obtained with each observation. The \( \varepsilon_{j(i)} \) are assumed normal with mean 0 and variance \( \sigma^2 \), and the subscript \( \varepsilon_{j(i)} \) indicates that the \( j \)th level of replication (pot) is nested within the \( i \)th level of treatment (different from notation ST&D). These terms (the treatment residuals) measure the variation within treatment groups. The new term \( \delta_{k(ij)} \) are the errors associated with each subsample (the pot residuals). It is convenient to think of the subsamples as being nested within each unique
combination of treatment and replication. The $\delta_k(ij)$ are also assumed normal with mean 0 and variance $\sigma^2$. We can rewrite the model this way:

$$Y_{ijk} = \bar{Y}_{i..} + (\bar{Y}_{i.} - \bar{Y}_{i..}) + (Y_{ij.} - \bar{Y}_{i.}) + (Y_{ijk} - \bar{Y}_{ij.}).$$

To get an intuitive sense of what this equation says, consider the plants within pots idea. $\tau_i$ measures the difference between a treatment mean and the overall mean (i.e. the treatment effect). $\epsilon_{j(i)}$ measures the difference between a pot mean and the mean of its assigned treatment (i.e. the experimental error, the variation among replications treated alike). $\delta_k(ij)$ measures the difference between a plant and the mean of its pot (i.e. the subsampling error).

3.5.2.2. Nested ANOVA with equal subsample numbers: computation

Following the example in ST&D, page 159.

In this experiment, mint plants are exposed to combinations of temperature and daylight and their one-week stem growth measured. The 6 treatment combinations (2 temperature levels by 3 light levels) are assigned randomly across 18 pots (i.e. 3 replications per treatment combination). Within each pot are four plants (i.e. subsamples).

Sometimes we may be uncertain as to whether a factor is crossed or nested. If the levels of the factor can be renumbered arbitrarily without affecting the analysis, then the factor is nested.

For example, pots 1,2,3 within treatment level 1 could be relabeled 2,3,1 without causing any problems. That is because pot number is simply an ID, not a classification variable. Pot 1 in treatment 1 has nothing to do with Pot 1 in treatment 2.

The data (from page 159):

<table>
<thead>
<tr>
<th>Plant N,</th>
<th>Low T, 8 hs</th>
<th>Low T, 12 hs</th>
<th>Low T, 16 hs</th>
<th>High T, 8 hs</th>
<th>High T, 12 hs</th>
<th>High T, 16 hs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pot number 1</td>
<td>Pot number 2</td>
<td>Pot number 3</td>
<td>Pot number 1</td>
<td>Pot number 2</td>
<td>Pot number 3</td>
</tr>
<tr>
<td>1</td>
<td>3.5 2.5 3.0</td>
<td>5.0 3.5 4.5</td>
<td>5.0 5.5 5.5</td>
<td>8.5 6.5 7.0</td>
<td>6.0 6.0 6.5</td>
<td>7.0 6.0 11.0</td>
</tr>
<tr>
<td>2</td>
<td>4.0 4.5 3.0</td>
<td>5.5 3.5 4.0</td>
<td>4.5 6.0 4.5</td>
<td>6.0 7.0 7.0</td>
<td>5.5 8.5 6.5</td>
<td>9.0 7.0 7.0</td>
</tr>
<tr>
<td>3</td>
<td>3.0 5.5 2.5</td>
<td>4.0 3.0 4.0</td>
<td>5.0 5.0 6.5</td>
<td>9.0 8.0 7.0</td>
<td>3.5 4.5 8.5</td>
<td>8.5 7.0 9.0</td>
</tr>
<tr>
<td>4</td>
<td>4.5 5.0 3.0</td>
<td>3.5 4.0 5.0</td>
<td>4.5 5.0 5.5</td>
<td>8.5 6.5 7.0</td>
<td>7.0 7.5 7.5</td>
<td>8.5 7.0 8.0</td>
</tr>
</tbody>
</table>

Pot totals = $Y_{ij}$
Treatment totals = $Y_{i.}$
Treatment means = $\bar{Y}_{i.}$

<table>
<thead>
<tr>
<th>Pot number</th>
<th>Low T, 8 hs</th>
<th>Low T, 12 hs</th>
<th>Low T, 16 hs</th>
<th>High T, 8 hs</th>
<th>High T, 12 hs</th>
<th>High T, 16 hs</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>17.5 11.5</td>
<td>18 14 17.5</td>
<td>19 21.5 22</td>
<td>32 28 28</td>
<td>22 26.5 29</td>
<td>33 27 35</td>
</tr>
<tr>
<td>44.0</td>
<td>49.5</td>
<td>62.5</td>
<td>88.0</td>
<td>77.5</td>
<td>95.0</td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>4.1</td>
<td>5.2</td>
<td>7.3</td>
<td>6.5</td>
<td>7.9</td>
<td></td>
</tr>
</tbody>
</table>

In this example, $t = 6$, $r = 3$, $s =$ number of subsamples = 4, and $n = trs = 72$. 

17
Recall that for a CRD the sums of squares satisfies:

$$\sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \bar{Y}_i)^2 = \sum_{i=1}^{t} (\overline{Y}_i - \overline{Y}_.)^2 + \sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \overline{Y}_i)^2 \text{ or } \text{TSS} = \text{SST} + \text{SSE}.$$ 

The degrees of freedom associated with these sums of square are n-1, t-1, and n-t, respectively. In the nested design, TSS and SST are unchanged but the SSE is partitioned into two components, the variation among pots within a treatment (experimental error) and the variation among plants within a pot (subsample error). The resulting equation can be written

$$\text{TSS} = \text{SST} + \text{SSEE} + \text{SSSE}$$

The two error terms represent the sum of squares due to experimental error and the sum of squares due to subsampling error.

**Nested ANOVA table:**

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Expected MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments (τ_i)</td>
<td>t - 1 = 5</td>
<td>SST</td>
<td>SST / 5</td>
<td>MST / MSEE</td>
<td>$\sigma_\delta^2 + 4\sigma_\varepsilon^2 + 12\Sigma \tau^2 / 5$</td>
</tr>
<tr>
<td>Exp. Error (ε_{ij(i)})</td>
<td>t (r - 1) = 12</td>
<td>SSEE</td>
<td>SSEE / 12</td>
<td>MSEE / MSSE</td>
<td>$\sigma_\varepsilon^2$</td>
</tr>
<tr>
<td>Samp. Error (δ_{k(ij)})</td>
<td>rt (s - 1) = 54</td>
<td>SSSE</td>
<td>SSSE / 54</td>
<td></td>
<td>$\sigma_\delta^2$</td>
</tr>
<tr>
<td>Total</td>
<td>trs - 1 = 71</td>
<td>TSS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In each case, the number of degrees of freedom is the product of the number of levels associated with each subscript between brackets and the number of levels minus one associated with the subscript outside the brackets.

In testing a hypothesis about treatment means, the appropriate divisor for $F$ is the mean square experimental error (MSEE) since it includes the variation from all sources (pot and plant) that contribute to the variability among treatment means except the treatment effects themselves.
Estimation of the different components of variance in the pot experiment

Again, the main objective of a nested design is to estimate the components of variance. To do this, we deconstruct the calculated mean squares according to their underlying theoretical models, called their expected mean squares (EMS, see last column in the above table) for each component of the linear model, as shown in the table below:

<table>
<thead>
<tr>
<th>Variance Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>Variance Component</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>71</td>
<td>255.91</td>
<td>3.60</td>
<td>4.05</td>
<td>100.0 %</td>
</tr>
<tr>
<td>Trtmt</td>
<td>5</td>
<td>179.64</td>
<td><strong>35.92</strong></td>
<td>2.81</td>
<td>69.4 %</td>
</tr>
<tr>
<td>Pot</td>
<td>12</td>
<td>25.83</td>
<td>2.15</td>
<td>0.30</td>
<td>7.5 %</td>
</tr>
<tr>
<td>Plant</td>
<td>54</td>
<td>40.43</td>
<td><strong>0.93</strong></td>
<td>0.93</td>
<td>23.0 %</td>
</tr>
</tbody>
</table>

\[ MSSE = \sigma_\delta^2 \]
\[ MSEE = \sigma_\delta^2 + 4\sigma_\epsilon^2 \]
\[ MST = \sigma_\delta^2 + 4\sigma_\epsilon^2 + 12\Sigma \tau^2/5 \]

so \( \sigma_\delta^2 = 0.93 \)
so \( \sigma_\epsilon^2 = (MSEE - \sigma_\delta^2)/4 = (2.15 - 0.93)/4 = 0.30 \)
so \( \Sigma \tau^2/5 = (MST - MSEE)/12 = (35.92 - 2.15)/12 = 2.81 \)

In this example, the variation among plants within a pot is three times larger than the variation among pots within a treatment.

In SAS, PROC VARCOMP computes these variance components for different models. For our example experiment here:

Proc GLM;
   Class Trtmt Pot;
   Model Growth = Trtmt Pot(Trtmt);
   Random Pot(Trtmt);
   Test h = Trtmt e = Pot(Trtmt);
Proc Varcomp;
   Class Trtmt Pot;
   Model Growth = Trtmt Pot(Trtmt);

**Pot(Trtmt):** This syntax tells SAS that the ID for pot only has meaning once the treatment level is specified. In other words, "Pot 1" does not tell you which pot in the experiment I'm talking about; but "Pot 1 in Treatment 1" does.

**Random Pot(Trtmt):** This syntax tells SAS that the pots are a random factor (i.e. pots 1, 2 and 3 are just a random sample from a larger population of pots which they represent).

**Test h = Trtmt e = Pot(Trtmt):** This syntax tells SAS which error term to use to test a particular hypothesis. For the hypothesis about treatments (h = Trtmt), the appropriate error term is the MSEE (i.e. Pot(trtmt)), so e = Pot(Trtmt). This specifies the test MST/MSEE.
You never specify the error term in a SAS model; the program knows there is always a source of error in the experiment. For a nested design, what this means is that you do not include a class variable for the lowest level of sub-sampling in the experiment (in this case, plant). By default, SAS will use the variation among plants within a pot as the error term for the experiment. This is why the test statement is so important: SAS's default error term (MSSE) is correct only for testing difference among pots within a treatment (MSEE), not for testing the differences among treatments (MST).

So, never include a class variable for the last level of sub-sampling; and always be sure to specify the correct error terms for your F tests.

If you have two levels of nesting (e.g. you measure two leaves from each plant), then you include plant as a class variable (but not leaf) and you indicate that a plant ID is only informative once you have specified both the pot and the treatment:

Proc GLM;
   Class Trtmt Pot Plant;
   Model Growth = Trtmt Pot(trtmt) Plant(Pot Trtmt);
   Random Pot(Trtmt) Plant(Pot Trtmt);
   Test h = Trtmt e = Pot(Trtmt);

Proc Varcomp;
   Class Trtmt Pot Plant;
   Model Growth = Trtmt Pot(trtmt) Plant(Pot Trtmt);

3.5.2.3. The optimal allocation of resources

(See Biometry Sokal & Rohlf page 309 for a detailed description). The main objective of a nested design is to investigate how the variation is distributed between among experimental units and among subsamples (i.e. where are the sources of error in the experiment). Once the variance component of the experimental units ($s_{e.u.}^2$) and the variance component of the subsamples ($s_{sub}^2$) are known, the variance of the means can be calculated:

$$s_Y^2 = \frac{s_{e.u.}^2}{n_{sub} \cdot r} + \frac{s_{sub}^2}{r}$$

Where $n_{sub}$ is the number of subsamples per experimental unit and $r$ is the number of replications per treatment. You can use this formula to test the effect of the different numbers of subsamples and replications on $s_Y^2$ (and thus the total information in the experiment) and use the different values to calculate relative efficiencies among designs.
However, the relative efficiency of one design with respect to another is not very meaningful unless the relative costs of the two designs are also taken into consideration. Clearly, if one design is twice as efficient as another but at the same time is ten times as expensive, we might not choose it. To introduce the idea of cost, we write a cost function. For a two-level nested design, the total cost (C) will be the cost of the subsamples multiplied by the total number of subsamples plus the cost of each experimental unit multiplied by the number of experimental units:

\[ C = n_{sub} \times C_{sub} + C_{eu} \]

To find the number of subsamples (n_{sub}) per experimental unit that will result in simultaneous minimal cost and minimal variance, the following formula may be used:

\[ n_{sub} = \sqrt{\frac{C_{e.u.} \times s_{e.u.}^2}{C_{sub} \times s_{sub}^2}} \]

The optimum number of subsamples will increase when the relative cost of the subsamples is low and the variance within experimental units is high (s_{e.u.}^2).

If the cost of samples and subsamples is the same, the optimum number of subsamples in our example can be calculated as:

\[ n_{sub} = \sqrt{\frac{s_{sub}^2}{s_{e.u.}^2}} = \sqrt{\frac{0.93}{0.30}} = 1.76 \text{ or } \approx 2 \text{ plants per pot} \]

If the cost is the same and s_{sub} < s_{e.u.}, it is better to allocate all the resources to experimental units (in this example, that means be put only one plant per pot). In terms of efficiency, subsampling is only useful when the variation among subsamples is larger than the variation among experimental units and/or the cost of the subsamples is smaller than the cost of the experimental units.