

## Topic 3. Single factor ANOVA: Introduction

[ST&D Ch. 7]

"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 [ST&D p. 99]

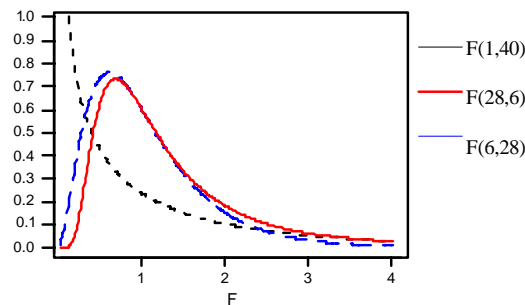


Fig 1. Three representative F-distributions. (Note  $F_{(1,40)}$  similar  $\chi^2_1$ )

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  $s^2_1$  (df:  $n_1 - 1$ ), followed by sampling  $n_2$  items and calculating their variance  $s^2_2$  (df:  $n_2 - 1$ ). We then calculate for each sample

$$F_s = s^2_1 / s^2_2$$

This ratio will be close to 1, because these variances are estimates of the same quantity. The expected distribution of this statistic is called **F-distribution**. The  $F$ -distribution is determined by **two** values for degrees of freedom. Statistical Tables for  $F$  show the cumulative probability distribution of  $F$  for several selected probability values. The values in the table represent  $F_{\alpha[v_1, v_2]}$  where  $\alpha$  is the proportion of the  $F$ -distribution to the right of the given-  $F$ -value (in one tail) and  $v_1, v_2$  are the degrees of freedom pertaining to the numerator and denominator of the variance ratio, respectively.

For example, a value  $F_{\alpha/2=0.025 [v_1=9, v_2=9]}=4.03$  indicates that the ratio  $s^2_1 / s^2_2$ , from samples of ten individuals from normally distributed populations with equal variance, is expected to be larger than 4.03 by chance in only **5%** of the experiments (the alternative hypothesis is  $s^2_1 \neq s^2_2$  so it is a **two tail test**)

### 3. 2. Testing the hypothesis of equality of variances [ST&D 116-118]

Suppose  $U_1, \dots, U_m$  are drawn from a normal distribution with mean  $m_u$  and variance  $s_u^2$ , and  $V_1, \dots, V_n$  are drawn from a normal distribution with mean  $m_v$  and

variance  $s_v^2$ . The F statistic can be used as a test for the hypothesis  $H_0: s_u^2 = s_v^2$  vs. the hypothesis  $H_A: s_u^2 \neq s_v^2$ .  $H_0$  is rejected at the  $\alpha$  level of significance if the ratio  $s_v^2 / s_u^2$  is either  $\geq F_{\alpha/2, m-1, n-1}$  or  $\leq F_{1-\alpha/2, m-1, n-1}$ . 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 [ST&D p. 98-112]

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

$$F = \frac{\text{estimate.of.}\sigma^2.\text{from.means}}{\text{estimate.of.}\sigma^2.\text{from.individuals}}$$

The denominator is an estimate of  $\sigma^2$  from the individuals in each sample. It is a **weighted average** of the sample variances.

The variance in 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  $n$ .

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

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

Varieties	Replications	$Y_{i.}$	$\bar{Y}_{i.}$	$s_{i.}^2$
1	19 14 15 17 20	85	$\bar{Y}_{1.} = 17$	6.5
2	23 19 19 21 18	100	$\bar{Y}_{2.} = 20$	4.0
		$Y_{..} = 185$	$\bar{Y}_{..} = 18.5$	

There are  $t = 2$  treatments and  $r = 5$  replications (the symbol  $t$  stands for "treatments" and  $r$  stands for "replications").

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

We will *assume* that the two populations have the same (unknown) variance  $s^2$  and test  $H_0: m_1 = m_2$ . We do this by obtaining two estimates for the variance  $s^2$  and comparing them.

First, we can compute the average sample variance **within samples**. To determine this variability called *experimental error*, we compute the variance of each sample ( $s_1^2$

and  $s^2_2$ ), assume they both estimate a common variance, and then estimate the common variance by pooling the sample variances.

$$s^2_1 = \frac{\sum (Y_{1j} - \bar{Y}_{1.})^2}{r-1} \quad s^2_2 = \frac{\sum (Y_{2j} - \bar{Y}_{2.})^2}{r-1} \quad \text{and}$$

$$s^2_w = \frac{(r_1 - 1)s^2_1 + (r_2 - 1)s^2_2}{(r_1 - 1) + (r_2 - 1)} = 4 * 6.5 + 4 * 4.0 / (4 + 4) = 5.25$$

Pooling  $s^2_1$  and  $s^2_2$  gives an estimate of  $s^2$  based on variability within the samples which we will designate  $s^2_w$  (subscript w= within).

The second estimate is based on the *between* samples variability. Assuming 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 estimate the variance of means using  $s^2_Y$ . Recall from Topic 1 that the mean  $\bar{Y}$  of a set of random variables drawn from a normal distribution with mean  $m$  and variance  $s^2$  is itself a normally distributed random variable with mean  $m$  and variance  $s^2/n$ .

The formula for  $s^2_Y$  is

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

and from the central limit theorem  $r$  times this quantity provides an estimate for  $s^2$  ( $r$  is the number of variates on which each sample mean is based).

Therefore the *between samples* estimate is

$$s_b^2 = r s^2_Y = 5 * 4.5 = 22.5$$

These two variances are used in the F test as follows. If the null hypothesis is not true, then the between sample variance should be much larger than the within sample variance. 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. Therefore we define the F statistic

$$F = s_b^2 / s_w^2$$

and test whether this statistic is significantly greater than 1. The F statistics measures how many times larger is the variability **between** the samples compared with the variability **within** the samples. For our experiment  $F = 22.5/5.25 = 4.29$ . The numerator  $s_b^2$  is based on 1 df, since there are two sample means. The denominator,  $s_w^2$ , is based on pooling the df within each sample so  $df s_w^2 = t(r-1) = 4 + 4 = 8$ . For these df we would expect an F value of 4.29 or larger just by chance about 7% of the times. From Table A.6,

p. 614 of ST&D,  $F_{0.05, 1, 8} = 5.32$ . 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

Since we have only two treatments and the square root of the F statistic is distributed according to a t distribution, and we can examine the t statistic:

$$F_{(1, n), 1 - \alpha} = t_{n, 1 - \alpha/2}^2 \qquad t = \sqrt{\frac{s_b^2}{s_w^2}}$$

The total degrees of freedom for the t statistic is  $t(r - 1) = tr - t$  since there are  $tr$  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  $t > t_{\alpha/2, tr-t}$ .

Here are the computations for our data set

The t statistic is the square root of this,  $t = 2.07$ . There are  $10 - 2 = 8$  degrees of freedom. From Table A.3, p. 611 of ST&D,  $t_{0.025, 8} = 2.306$  (note that  $2.306^2 = 5.32 = F_{0.05, 1, 8}$ ). Therefore we fail to reject  $H_0$  at the 0.05 significance level.

## 3. 4. The linear additive model [ST&D p. 32, 103, 152]

**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  $\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i$  (epsilon sub i) is applicable to the problem of estimating or making inferences about population means and variances. This model attempts to explain an observation as a mean  $\boldsymbol{\mu}$  plus a random element of variation  $\boldsymbol{\varepsilon}_i$ . The  $\boldsymbol{\varepsilon}_i$ 's are assumed to be from a population of uncorrelated  $\boldsymbol{\varepsilon}$ 's with mean zero. Independence among  $\boldsymbol{\varepsilon}$ 's is assured by random sampling.

**3. 4. 2. Two populations:** This model is more general than the model of 3. 4. 1. because it permits us to describe two populations simultaneously. For samples from **two** populations with possibly different means but a common variance, the composition of any observation is given by  $\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\varepsilon}_{ij}$ . Therefore a given reading is composed of the grand mean  $\boldsymbol{\mu}$  of the population, a component  $\boldsymbol{\tau}_i$  for the population involved ( $\boldsymbol{\mu} + \boldsymbol{\tau}_1 = \boldsymbol{\mu}_1$  and  $\boldsymbol{\mu} + \boldsymbol{\tau}_2 = \boldsymbol{\mu}_2$ ), and a random deviation  $\boldsymbol{\varepsilon}_{ij}$ . The subindexes  $i = 1$  and  $2$  indicate the treatment number and the subindexes  $j = 1, \dots, r$  indicate the number in each sample (replications).

The  $\boldsymbol{\tau}_i$  are measured as deviations from a mean that is set as a reference point as  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) / 2$  so that  $\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 = \mathbf{0}$  or  $-\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2$ . This does not affect the difference between means that is  $2\boldsymbol{\tau}$ . If  $n_1 \neq n_2$  we may set  $n_1\boldsymbol{\tau}_1 + n_2\boldsymbol{\tau}_2 = 0$ .

The  $\boldsymbol{\varepsilon}$ 's are assumed to be from a single population with normal distribution, mean  $\boldsymbol{\mu} = 0$ , and variance  $s^2$ .

The data represent the model as:

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

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

As with the 2 sample t-test, the linear model is written  $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$  where  $i = 1, \dots, t$  and  $j = 1, \dots, r$ . 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  $\tau$ 's that will differentiate the **Model I ANOVA** from the **Model II ANOVA**.

- **The Model I ANOVA or fixed model:** In this model the  $\tau$ 's are fixed and

$$\sum \tau_i = 0$$

Setting the  $\sum \tau_i = 0$  is simply measuring treatment effects as deviations from an overall mean. The null hypothesis is then stated as  $H_0$ : is  $\tau_1 = \dots = \tau_t = 0$  and the alternative as  $H_1$ : some is some  $\tau_i \neq 0$ . What Model I anova tests is the **differential** effects of the 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  $\tau_i$ . The  $\tau$ 's constitute a finite population and are the parameters of interest along with  $s^2$ . When the null hypothesis is false and some  $\tau_i \neq 0$  there will be an additional component due to treatment effects equal to:

$$r \sum \tau_i^2 / (t-1)$$

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 added effects for each group ( $\tau$ 's) are not fixed treatments but are random effects. 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  $t$ 's are a random sample from a population of  $\tau$ 's for which the mean is zero and the variance is  $\sigma^2_t$ . When the null hypothesis is false there will be an additional component of variance equal to  $r\sigma^2_t$ . 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:  $\sigma^2_t$ . We test for its presence and estimate its magnitude, as well as its percentage contribution to the variation. For the fixed model, we draw inferences about the particular treatments; for the random model, we draw an inference about the population of treatments. The null hypothesis 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. The purposes differ for the two models, as do some of the supplementary tests and computations following the initial significance test. At present we will deal only with the **fixed model**.

- **Assumptions of the model** [ST&D p. 174]
  1. Treatment and environmental effects are additive

2. Experimental errors are random, independently and normally distributed about zero mean and with a common variance

Error terms are independently and normally distributed

There is no relation between the experimental grouping of treatments and the size of the error terms. This could be violated if plots were not chosen randomly.

Variances are homogeneous

The variances of different treatments are the same.

Means and variances are independent

For example, suppose yield is measured and the treatment causes yield to range from 1 gm/ plant up to 10 gm/plant. A range of  $\pm 1$  gm would be much more "significant" at the low end than the high end but could not be considered any differently.

Effects are additive

Random error term is added rather than being, for example, multiplied.

### 3. 5. ANOVA. Single factor designs

#### 3. 5. 1. The Completely Random Design CRD

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

Table 1. TREATMENT (Inoculation of Rhizobium strains, Table 7.1)

	3DOK1	3DOK5	3DOK4	3DOK7	3DOK13	composite	Total
	19.4	17.7	17.0	20.7	14.3	17.3	
	32.6	24.8	19.4	21.0	14.4	19.4	
	27.0	27.9	9.1	20.5	11.8	19.1	
	32.1	25.2	11.9	18.8	11.6	16.9	
	33.0	24.3	15.8	18.6	14.2	20.8	
$\Sigma Y_{ij} = Y_{i.}$	144.1	119.9	73.2	99.6	66.3	93.5	596.6 = $Y_{..}$
$\Sigma Y_{ij}^2$	4287.53	2932.27	1139.42	1989.14	887.29	1758.71	12994.36
$Y_{i.}^2/r$	4152.96	2875.2	1071.65	1984.03	879.14	1748.45	12711.43
$\Sigma (Y_{ij} - \bar{Y}_{i.})^2$	134.57	57.07	67.77	5.11	8.15	10.26	282.93
$\bar{Y}_{i.} = \text{mean}$	28.8	24.0	14.6	19.9	13.3	18.7	19.88
$\sigma_{n-1}^2$ variance	33.64	14.27	16.94	1.28	2.04	2.56	

The completely randomized design is the basic ANOVA design. It is used when there are  $t$  different treatment levels (treatments) of a single factor (e.g. clover culture). These treatments are applied to  $t$  independent random samples of size  $r$  (NOTE:  $r$  can depend on the treatment number, but the formulas become much more confusing and we will postpone them till later). The total sample size is  $n = rt$ . Let  $Y_{ij}$  denote the  $j^{\text{th}}$

measurement recorded from the  $i^{\text{th}}$  treatment. WARNING: Some texts interchange the  $i$  and the  $j$  (i.e. the rows and columns of the table)!

We wish to test the hypothesis  $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_t$  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

$$F = s_b^2 / s_w^2$$

Here is another way to write this. In our new dot notation we can write

$$s_w^2 = \frac{\sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{i.})^2}{t(r-1)} = \frac{SSE}{t(r-1)} \quad \text{where } SSE = \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{i.})^2$$

Here SSE is the *sum of squares for error*. Also

$$s_b^2 = \frac{r \sum_{i=1}^t (\bar{Y}_{i.} - \bar{Y}_{..})^2}{t-1} = \frac{SST}{t-1} \quad \text{where } SST = r \sum_{i=1}^t (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

Here SST is the *sum of squares for treatments* (Model Sum of Squares in SAS). Since, variances among means estimate  $\sigma^2/r$ , the  $r$  in the definition formula for treatment SS assures us that the MST estimates  $\sigma^2$  rather than  $\sigma^2/r$ . In the example 3.3 we also multiplied by  $r$  to estimate the *between samples* variance ( $s_b^2 = r s_y^2$ )

In this notation we can write

$$F = \frac{SST/(t-1)}{SSE/t(r-1)}$$

Since  $rt = n$ , we can also write

$$F = \frac{SST/(t-1)}{SSE/(n-t)}$$

We can then define

**MSE** =  $SSE/(n-t)$ , the *mean square for error* it gives the average dispersion of the items around the group means. It is an estimate of a common  $\sigma^2$ , the within variation or 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.

**MST** =  $SST/(t-1)$ , the *mean square for treatments* (MS Model in SAS) is an independent estimate of  $\sigma^2$ , when the null hypothesis is true ( $H_0: m_1 = m_2 = m_3 = \dots = m_t$ ). If there are differences among treatment means there will be an added component due to

treatment effects equal to  $r\sum\tau_i^2/(t-1)$  (Model I) or  $r\sigma_t^2$  (Model II) (see topic 3.4.3. and ST&D p. 155).

### **$F = MST/MSE$**

The  $F$  value is obtained by dividing the treatment mean square by the error mean square. We expect to find  $F$  approximately equal to  $\sigma^2/\sigma^2=1$ . In fact however the expected ratio is

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

It is clear from this formula, that the  $F$ -test is sensitive to the presence of the added component due to treatment effects. The ANOVA permits us to test whether there are added 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 number of degrees of freedom is the number of independent quantities in the statistic.

Thus SST has the  $t$  quantities  $(\bar{Y}_i - \bar{Y}_..)$  which have one constraint that they must sum to 0, and SSE has the  $n$  quantities  $Y_{ij}$ , which have  $t$  constraints for the  $t$  sample means.

We can also use the equation

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

If you look at the above formula carefully, you will see that it is somewhat surprising. There is a cross product having terms of the form  $2(\bar{Y}_{ij} \bar{Y}_..)$  that should appear. It turns out that all of these cross product terms cancel each other out so the sum is zero. Quantities that satisfy this 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 portion due to variation within groups. The degrees of freedom are also additive.

The actual calculations, when done by hand, use the formulas

$C = (Y_{..})^2 / n = (\sum_{ij} Y_{ij})^2 / n$	The correction term. Is the squared sum of all observations divided by their number.
$SS = \sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - C$	The sum of squares that includes all sources of variation. This is the total SS.
$SST = \sum_{i=1}^t Y_{i.}^2 / r - C$	The sum of squares attributable to the variable of classification. This is the between SS, or among groups SS or treatment SS.
$SSE = SS - SST$	The sum of squares among individuals treated alike. This is the within groups SS, or residual SS or error SS.

This information can be summarized in an ANOVA table:

Source	df	Definition	SS	MS	F
Treatments	t - 1	$r \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$	SST	SST/(t-1)	MST/MSE
Error	t(r-1) = n - t	$\sum_{i,j} (Y_{ij} - \bar{Y}_{i.})^2$	SS - SST	SSE/(n-t)	
Total	n - 1	$\sum_{i,j} (Y_{ij} - \bar{Y}_{..})^2$	SS		

Table 2. One way ANOVA (Table 7.3, page 144 ST&D)

Source	df	SS	MS	F
Among cultures	5	847.05	169.41	14.37**
Within cultures	24	282.93	11.79	
Total	29	1129.98		

Notice that, MSE (11.79) is the pooled variance or the average of treatment variances, i.e.,  $MSE = \sum s_i^2 / k$ ; where  $s_i^2$  is the variance estimated from the  $i$ th treatment. The  $F$  value of 14 indicates that the variation between treatments is 14 times larger than the normal variation within treatments.

### 3. 5. 1. 2. Assumptions associated with ANOVA

The assumptions associated with ANOVA can be expressed in terms of a statistical model as follows. The measurement  $Y_{ij}$  is written

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}.$$

Then  $\varepsilon_{ij}$  is assumed to be normally distributed with mean 0 and variance  $s^2$  (independent of treatment level  $i$  and sample number  $j$ ).

#### 3. 5. 1. 2. 1. Normal distribution:

The Shapiro and Wilk test statistics  $W$  (ST&D p.567, and SAS PROC UNIVARIATE NORMAL; or Analyst Application) is a powerful test for normality for small to medium samples ( $n < 2000$ ). Normality is rejected if  $W$  is sufficiently smaller than 1.  $W$  is similar to a correlation between the data and their normal scores (ST&D 566). In a perfectly normal population there is a perfect correlation  $W=1$ .

For large populations ( $n > 2000$ ) SAS recommends the use of the Kolmogorov-Smirnov statistics (ST&D 571).

#### 3. 5. 1. 2. 2. Homogeneity of variances:

The Tests for Equal Variance 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 more robust to deviations from normality.

Basically Levene's test is an ANOVA of the squares of the residuals of each observation from the treatment means. If Levene's test rejects the hypothesis of homogeneity of variances *Analyst Application* provides the option to perform 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 need to use the option HOVTEST (for Homogeneity of variance test) within the means statement in the PROC GLM procedure.

```
proc glm;
  class Tr;
  model De = Tr;
  means Tr / Hovtest= Levene;
run;
```

### 3. 5. 1. 3. Experimental Procedure

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

Here is how the clover plots might look if this experiment were conducted in the field. The experimental procedure would be: First randomly (e.g. from a random number table such as ST&D p. 606, or using PROC PLAN in SAS) select the block numbers for treatments A,B,C,D,E, and F. Take for treatment A the first 5 random numbers under 30, and so forth. Example: on p. 607, starting from row 02, column 88-89 (these numbers picked by pointing at locations in the table) and move downward: Treatment A: 5,19,13, 20, 6; B: 14, 26,1, 8, 4; etc. Then take the measurements. These are given in the table.

### 3. 5. 1. 4. Power and sample size.

Pearson and Hartley (1953, Biometrika 38:112-130) provided power function charts that are easy to use to calculate the power of an ANOVA and the appropriate number of replications.

#### 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 for the ANOVA it is necessary to calculate first the critical value  $\phi$ . This critical value depends on the number of treatments (k), the number of replications (n), the magnitude of the treatment effects that the investigator wishes to detect (d), an estimate of the population variance ( $\sigma^2 = MS_{\text{error}}$ ), and the probability of rejecting a true null hypothesis ( $\alpha$ ).

In a CRD  $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$  ;  $i = 1, 2, ..k$ ;  $j = 1, 2 \dots n$

where  $\mu$  is the overall mean, and  $\tau_i$  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:

$$\phi = \sqrt{\frac{r}{MSE} \sum \frac{\tau_i^2}{k}}$$

This general formula can be **simplified** if we assume all  $\tau_i$  are zero except the two extreme treatment effects which are  $\mu_{(k)}$ ,  $\mu_{(1)}$  and  $d = \mu_{(k)} - \mu_{(1)}$ . You can think

of  $d$  as the difference between the extreme means. Since  $u$  is in the middle of  $\mu_{(k)}$  and  $\mu_{(1)}$ ,  $\tau_i = d/2$

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

And the  $\phi$  formula simplifies to  $\phi = \sqrt{\frac{d^2 * r}{2k * MS_{error}}}$

Entering the chart for  $v_1 = df = k-1$  and the section for the appropriate  $\alpha$  (0.05 or 0.01), the interception of the calculated  $\phi$  and  $v_2 = df = k(n-1)$  gives the power of the test (both sides of the chart).

**Example:** Suppose that one experiment has  $k=6$  treatments with  $r=2$  replications each. The difference between the extreme means was 10 units,  $MSE= 5.46$ , and the required  $\alpha = 5\%$ . To calculate the power:

$$\phi = \text{SQR}(10^2 * 2 / 2 * 6 * 5.46) = 1.75$$

Use Chart  $v_1 = k-1 = 5$ . Use the set of curves to the left ( $\alpha = 5\%$ ). Select curve  $v_2 = k(r-1) = 6$ . 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.

To calculate the power using the *Analyst*: Statistics → ANOVA → One-Way ANOVA → Tests → Power analysis. Or *Analyst* → Sample Size → One-Way ANOVA → 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 ‘n’ for a given  $\alpha$  and power: a) specify the constants, b) start with an arbitrary number of ‘n’ to compute  $\phi$ , c) use Pearson and Hartley’s charts to find the power, and d) iterate the process until a minimum ‘n’ value which satisfies a required power for a given  $\alpha$  level is found.

**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,  $k = 6$ ,  $\alpha = 1\%$ ,  $\beta = 10\%$ ,  $d = 15$  and  $\sigma^2 = 12$ ).

n	df	$\phi$	(1- $\beta$ ) for $\alpha=1\%$
2	6(2-1)= 6	1.77	0.22
3	6(3-1)= 12	2.17	0.71
4	6(4-1)= 18	2.50	0.93

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

---

### 3. 5. 2. Subsampling: the nested design [ST&D p.157 - 167]

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 made on subsamples or *sampling units*. The classical example of this is that 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, etc. We call the analysis of this kind of data organized in a hierarchical way *nested analysis of variance*. Nested anovas are not limited to two levels, we can divide the subgroups into sub-subgroups, and even further, as long as these are chosen randomly.

The applications of nested anova are:

- To ascertain the magnitude of error at various stages of an experiment or process
- To estimate the magnitude of the variance attributable to various levels of variation in a study of quantitative genetics
- To discover sources of variation in natural population in systematic studies, etc.

#### 3. 5. 2. 1. Linear model for subsampling:

Before we compute an actual nested anova, we should examine the linear model upon which it is based:

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

The interpretations of  $\mu$ ,  $\tau$ , and  $\varepsilon$  are as before. Two random elements are obtained with each observation. The  $\varepsilon_{j(i)}$  are assumed normal with mean 0 and variance  $\sigma_\varepsilon^2$ , the subscript  $\varepsilon_{j(i)}$  indicates that the  $j^{\text{th}}$  level of factor B (pot) is nested under the  $i^{\text{th}}$  level of factor A (treatment) (different from notation ST&D). They measure the treatment error. The  $\delta_{k(ij)}$  are the error associated with each subsample. It is convenient to think of the replicates as being nested within the combination of levels of A and B. The  $\delta_{k(ij)}$  are assumed normal with mean 0 and variance  $\sigma^2$ . This is represented in the data as

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

The dot notation: the dot replaces a subscript and indicates that all values covered by that subscript have been added

### 3. 5. 2. 2. Nested ANOVA with equal subsample numbers: computation

ST&D give an example on p. 159. There are 6 treatments (combinations of temperature and daylight) and 3 pots nested under each level of treatment (replications), and 4 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, as for example Pot N<sup>o</sup>: 1, 2, 3, 4, 5, 6, ..., 17, 18, then the factor is nested. Since every level of the nested factor does not appear with every level of the treatment factor, there can be no interaction between these two factors. The data look like this:

	Low T, 8 hs	Low T, 12 hs	Low T, 16 hs	High T, 8 hs	High T, 12 hs	High T, 16 hs
Plant N <sub>o</sub>	Pot number 1 2 3	Pot number 1 2 3	Pot number 1 2 3	Pot number 1 2 3	Pot number 1 2 3	Pot number 1 2 3
1	3.5 2.5 3.0	5.0 3.5 4.5	5.0 5.5 5.5	8.5 6.5 7.0	6.0 6.0 6.5	7.0 6.0 11.0
2	4.0 4.5 3.0	5.5 3.5 4.0	4.5 6.0 4.5	6.0 7.0 7.0	5.5 8.5 6.5	9.0 7.0 7.0
3	3.0 5.5 2.5	4.0 3.0 4.0	5.0 5.0 6.5	9.0 8.0 7.0	3.5 4.5 8.5	8.5 7.0 9.0
4	4.5 5.0 3.0	3.5 4.0 5.0	4.5 5.0 5.5	8.5 6.5 7.0	7.0 7.5 7.5	8.5 7.0 8.0
Pot totals = Y <sub>ij</sub> .	15 17.5 11.5	18 14 17.5	19 21.5 22	32 28 28	22 26.5 29	33 27 35
Treatment totals = Y <sub>i..</sub>	44.0	49.5	62.5	88.0	77.5	95.0
Treatment means = $\bar{Y}_{i..}$	3.7	4.1	5.2	7.3	6.5	7.9

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

Recall that for the CRD the sums of squares satisfies

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

The degrees of freedom associated with these sums of square are  $n-1$ ,  $t-1$ , and  $n-t$ , respectively. In the nested design the SST is unchanged but the SSE is further partitioned into two components. The resulting equation can be written

$$\sum_{i=1}^t \sum_{j=1}^r \sum_{k=1}^s (Y_{ijk} - \bar{Y}_{...})^2 = rs \sum_{i=1}^t (\bar{Y}_{i..} - \bar{Y}_{...})^2 + s \sum_{i=1}^t \sum_{j=1}^r (Y_{ij.} - \bar{Y}_{i..})^2 + \sum_{k=1}^s (\bar{Y}_{ijk} - \bar{Y}_{ij.})^2$$

or  $\mathbf{SS} = \mathbf{SST} + \mathbf{SSEE} + \mathbf{SSSE}$

The two error terms represent the sum of squares due to *experimental error* and the sum of squares due to *sampling error*. The SSEE represents the effects of the  $e_{j(i)}$  term, which is the variation among plots (pots) within treatments. The SSSE represents the variation among subsamples (plants) within plots (pots).

**ANOVA table:**

Source of variation	df	SS	MS	F	Expected MS
Treatments ( $t_i$ )	$t - 1 = 5$	SST	SST/ 5	<b>MST/MSEE</b>	$\sigma_\delta^2 + 4\sigma_\varepsilon^2 + 12\Sigma\tau^2/5$
Exp. Error ( $e_{j(i)}$ )	$t(r - 1) = 12$	SSEE	SSEE/ 12	MSEE/MSSE	$\sigma_\delta^2 + 4\sigma_\varepsilon^2$
Samp. Error ( $d_{k(ij)}$ )	$rt(s - 1) = 54$	SSSE	SSSE/ 54		$\sigma_\delta^2$
Total	$trs - 1 = 71$	SS			

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 population treatment means, the appropriate divisor for  $F$  is the experimental error MS since it includes variation from all sources that contribute to the variability of treatment means except treatments.

**Estimation of the different components of variance in the pot experiment**

The main objective in a nested design is to estimate the variance components:

Variance Source	df	Sum of Squares	Mean Squares	Variance component	Percent of total
Total	71	255.91	3.60	4.05	100.0 %
trtmt	5	179.64	35.92	2.81	69.4 %
pot	12	25.83	2.15	0.30	7.5 %
error	54	40.43	0.93	0.93	23.0 %

$$MSSE = \sigma_\delta^2$$

$$\sigma_\delta^2 = 0.93$$

$$MSEE = \sigma_\delta^2 + 4\sigma_\varepsilon^2$$

$$\sigma_\varepsilon^2 = (MSEE - \sigma_\delta^2)/4 = (2.15 - 0.93)/4 = 0.30$$

$$MST = \sigma_\delta^2 + 4\sigma_\varepsilon^2 + Tr$$

$$Tr = (MST - MSEE)/12 = (35.92 - 2.15)/12 = 2.81$$

The expected Mean Squares (last column on the previous ANOVA Table) are used to calculate the different variance components of the model. In this example the variation among plants within a pot is three times larger than the variation among pots.

**SAS PROC VARCOMP** provides a way to do the computation of these variance components in different models. For the pots experiment:

SAS CODE

```
proc GLM;
  class trtmt pot;
  model growth= trtmt pot (trtmt);
  random pot (trmt);
  test h=trmt e=pot(trtmt);
proc varcomp;
  class trtmt pot;
  model growth= trtmt pot(trtmt);
run; quit;
```

**pot (trtmt):** indicates that pot is a nested factor in treatment. Pot 1 in treatment 1 is not more similar to pot 1 in treatment 2 than to pots 2 and 3.

**Random pot (trtmt):** indicates that the pots are a random factor. Pots 1, 2 and 3 are just a random sample of pots.

**test h=trmt e=pot(trtmt):** indicates which error term to use to test a particular hypothesis. For the hypothesis about treatments (**h=trmt**) use pot(trtmt) as the error term (**e=pot(trtmt)**; this is the same as **MST/MSEE**)

Note that you never include a class variable for the last level of sub-sampling (in this case plant).

If you have **two levels of nesting**, e.g. you take two samples from each plant then you include plant as a class variable (but not sample) and you indicate that plant is nested in (pot trtmt)

```
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);
run; quit;
```

### 3.5.2.3. The optimal allocation of resources

(Additional information in: Biometry Sokal & Rohlf pg. 309 for a detailed description)

One of the main reasons to do a nested design is to investigate how is the variation distributed between experimental units and subsamples. Once the variance of the experimental units ( $s_{e.u.}^2$ ) and the variance of the subsamples  $s_{SUB}^2$  are known, the variance of the means can be calculated as

$$s_Y^2 = \frac{s_{eu}^2}{N_s * N_r} + \frac{s_{SUB}^2}{N_r}$$

Where  $N_s$  is the number of subsamples per experimental unit and  $N_r$  is the number of replications. You can use this formula to test the effect of the different numbers of subsamples and replications in the  $s_Y^2$ , and use the different values to calculate the relative efficiency.

However the relative efficiency of one design with respect to another is not very meaningful, unless the relative cost of obtaining the two designs are 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 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:

$$C = Ns * Nr(C_{SUB}) + Nr(C_{eu})$$

To calculate the number of subsamples ( $Ns$ ) per experimental unit that will result in simultaneous minimal cost and minimal variance the following formula may be used:

$$Ns = \sqrt{\frac{C_{e.u.} * S_{SUB}^2}{C_{SUB} * S_{e.u.}^2}}$$

The optimum number of subsamples will increase when the relative cost of the subsamples is low and the variance within the experimental unit is high ( $S_{SUB}^2$ ).

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

$$Ns = \sqrt{\frac{S_{SUB}^2}{S_{e.u.}^2}} = \sqrt{\frac{0.93}{0.30}} = 1.76 \text{ or } \approx 2 \text{ plants per pot}$$

If the cost is the same and the  $s_{sub} < s_{e.u.}$ , it is better to allocate all the replications to the experimental units (in this example would be to put one plant per pot). Therefore 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.