BIOSTATISTICS
TOPIC 7: ANALYSIS OF DIFFERENCES
II. MULTIPLE COMPARISONS

In Topic 6, we discussed methods for comparing two groups. However, many clinical experiments involve more than two treatments. In this topic, we discuss methods for comparing g ($g > 2$) treatment groups, where the treatments are randomly assigned to patients. For example, a clinical trial might be interested in comparing the efficacy of 5 drugs in relation of improvement in bone mineral density. It would seem this problem could be solved by performing a t-test on all possible pairs of means. However, this solution would be incorrect, since it leads to considerable distortion of a statistical *type I error*. For instance, in the above example, there are 10 possible pairs and if the probability of correctly accepting the null hypothesis for *each* pair comparisons is 1 - 0.05 = 0.95, then the probability of correctly accepting the null hypothesis for *all* 10 tests is $(0.95)^{10} = 0.60$, if the tests are independent. Thus, a substantial increase in the type I error has occurred.

The appropriate procedure for testing the equality of several means is the analysis of variance (ANOVA). However, ANOVA has a wider application than the problem above. It is probably the most useful technique in the field of statistical inference. The topic is an extensive subject to which numerous books have entirely devoted to the subject because it is directly linked to the issues of design of experiments. The problem of design is, of course, inseparable from those of analysis and it is worth emphasizing that unless a sensible design is employed, it may be very difficult or even impossible to obtain valid conclusions from the resulting data. Before studying the ANOVA technique, let us discuss the concept of effect and replication.

I. THE CONCEPT OF EFFECT AND LINEAR MODEL

(A) GENERAL INTRODUCTION

Let us start with a simple example: suppose that we have three samples, each with three observations, represent identical population distributions, and that there is *no*

variability (that is, no error) within any of the populations. If the mean of each of the populations is $\mu = 40$, then our sample results should perhaps look like this:

Sample 1	Sample 2	Sample 3
40	40	40
40	40	40
40	40	40

There should be no differences either between or within samples if this is the true situation. When this is true and let the observation for each individual *i* in each group *j* be y_{ij}, we could write:

$$y_{ij} = \mu \ ,$$

where μ is, of course, a constant (μ = 40). Now suppose that the three samples are given different treatments, and that treatments produce effects, but that there is once again no variability within a treatment population (again, no error). Our results might look like:

Sample 1	Sample 2	Sample 3
40 - 2 = 38	40 + 6 = 46	40 - 4 = 36
40 - 2 = 38	40 + 6 = 46	40 - 4 = 36
40 - 2 = 38	40 + 6 = 46	40 - 4 = 36

Here there are differences between observations in different treatments, but there are no differences within a treatment sample. The linear model here is:

$$y_{ij} = \mu + \alpha_j$$

where, as can be seen, $\alpha_1 = -2$, $\alpha_2 = 6$ and $\alpha_3 = -4$. Note that the sum of treatment effect is zero.

In reality, there is always variability in a population, so that there is sampling error. The actual data we might obtain would undoubtedly look something like:

	Sample 1	Sample 2	Sample 3
	40 - 2 + 5 = 43	40 + 6 - 5 = 41	40 - 4 + 3 = 39
	40 - 2 + 2 = 40	40 + 6 + 1 = 47	40 - 4 - 2 = 34
	40 - 2 - 3 = 35	40 + 6 + 8 = 654	40 - 4 + 1 = 37
Mean	39.3	47.3	36.7

Overall mean: 41.1

Here a random error component has been added to the value of μ and the value of α_j in the formation of each score. The linear model in this situation is then:

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

Notice that not only do differences exist between observations in different treatments, but also between observations in the same treatment.

If we estimate the effect of treatment 1 by taking

$$\text{est } \alpha_1 = \bar{x}_1 - \bar{x} = 39.3 - 41.1 = -1.8$$

it happens that we are almost right, since the data were simulated so that $\alpha_1 = -2$. Likewise, our estimate of α_2 is in error by 0.2 and our estimate of α_3 is error by -0.4.

This example is to point out that evidence for experimental effects has something to do with the differences *between* the different groups relative to the differences that exist *within* each group. Next, we will turn to the problem of partition the variability among observations into two parts: the part that should reflect both experimental effects and sampling error, and the part that should reflect sampling error alone.

(B) PARTITIONING OF VARIANCES

We begin by denoting the observation from an individual *i* belong to sample *j* be y_{ij}, the overall mean by $\bar{y}$ and the mean for each sample *j* by $\bar{y}_j$, then we could write:

$$y_{ij} - \bar{y} = (y_{ij} - \bar{y}_j) + (\bar{y}_j - \bar{y})$$

On squaring both sides of this equation we obtain:

$$\sum_j \sum_i (y_{ij} - \bar{y})^2 = \sum_j \sum_i [(y_{ij} - \bar{y}_j) + (\bar{y}_j - \bar{y})]^2$$

$$= \sum_j \sum_i (y_{ij} - \bar{y}_j)^2 + \sum_j \sum_i (\bar{y}_j - \bar{y})^2 - 2\sum_j \sum_i (y_{ij} - \bar{y}_j)(\bar{y}_j - \bar{y})$$

The expression is a little bit complicated. Now, let us analyse one by one: first, notice that the term $2\sum_j \sum_i (y_{ij} - \bar{y}_j)(\bar{y}_j - \bar{y})$ is first to be summed over *i* and then over *j*. But $(\bar{y}_j - \bar{y})$ is the same for all *i* in the *j* sample and the sum of $(y_{ij} - \bar{y}_j)$ is zero, therefore:

$$2\sum_j \sum_i (y_{ij} - \bar{y}_j)(\bar{y}_j - \bar{y}) = 2\sum_j (\bar{y}_j - \bar{y}) \sum_i (y_{ij} - \bar{y}_j) = 0$$

Second, the term $\sum_j \sum_i (\bar{y}_j - \bar{y})^2$ is essentially the deviation between each sample mean and overall mean. Again, the sum is over all *i* and then over all *j*. If the sample size of each sample (group) is n_j, then the sum $\sum_{i=1}^{n_j} (\bar{y}_j - \bar{y})^2$ i.e. over n_j times. In other words:

$$\sum_j \sum_i (\bar{y}_j - \bar{y})^2 = \sum_j n_j (\bar{y}_j - \bar{y})^2$$

Finally, the term $\sum_j \sum_i (y_{ij} - \bar{y}_j)^2$ could be obtained as the differences between $\sum_j \sum_i (y_{ij} - \bar{y})^2$ and $\sum_j \sum_i (\bar{y}_j - \bar{y})^2$.

That is, the total sum squares (SS) can be written as:

$$\sum_j \sum_i (y_{ij} - \bar{y})^2 = \sum_j \sum_i (y_{ij} - \bar{y}_j)^2 + \sum_j n_j (\bar{y}_j - \bar{y})^2$$

Toal SS = Within SS + Between SS

II. WHAT IS REPLICATION ?

Consider the following experiment to compare two treatments applied to only two patients. Suppose that treatment A1 gives a response of 180 units and treatment B1 gives 168 units. Then we would suspect that treatment A was better than treatment B. But, we have no idea if the difference of (180 - 168) = 12 is due to treatment *effect* or due to the natural variability. Even if there is no treatment effect, it is highly unlikely that the results will be exactly the same.

Now suppose that the experiment is repeated on a next two patients and that the following results are obtained. A2 = 176 and B2 = 171. Then an estimate of the treatment effect is obviously (A1 + A2 - B1 - B2) / 2 = 8.5.

But now, we have also two estimates of residual variation, namely, (A1-A2) and (B1-B2). These can be combined in two ways to give:

(A1 - A2 + B1 - B2) = 0.5
(A1 - A2 - B1 + B2) = 3.5

The treatment effect (8.5) is much larger than the other two comparisons (0.5 and 3.5) and this is a definite indication that treatment A is better than treatment B. Since there are only two groups, we can use the t-test (topic 6); it can be shown that the estimate of the residual standard deviation is given by:

$$s = \sqrt{\frac{(A1 - A2)^2 + (B1 - B2)^2}{4}} = 2.5$$

and the standard error of the estimate treatment effect (difference) is:

$$SE(Diff) = s\sqrt{\frac{1}{2}+\frac{1}{2}} = 2.5$$

Thus the value of the standardised distance is:

$$t = 8.5 / 2.5 = 3.4$$

which is actually less than its expected value (t = 4.3 with 2 df and alpha=5%). In other words, the result is not statistically significant. It would be advisable to make more observations (or replications) in order to improve the power of the test. The process of design and analysis of a controlled experiment in which several replications are made in one treatment will be considered in the context of analysis of variance as follows:

III. SINGLE FACTOR (ONE-WAY) ANALYSIS OF VARIANCE

It is perhaps best to start this subject with a concrete example as follows:

Example 1: The weight gain in pounds over three weeks of 35 pigs from five different treatments are given in the following table:

	Treatment				
	1	2	3	4	5
	23	29	38	30	31
	27	25	31	27	33
	26	33	28	28	31
	19	36	35	22	28
	30	32	33	33	30
	30	28	36	34	24
	27	30	35	34	29
	25	31	37	32	30
Number of pigs:	8	8	8	8	8
Sum:	207	244	273	240	236
Mean:	28.75	30.5	34.13	30	29.5
Variance	13.26	11.14	10.98	17.43	7.14

It was interested to know whether the weight gains were different between treatment groups?

2.1. PARTITION OF VARIATIONS

Let us denote the weight gain for an ith treatment in a jth pig be x_{ij}, that is, i = 1,2,..,5 and j = 1,2,...,35. Then, the overall mean $\bar{x}$ is estimated by:

$$\bar{x} = \sum_{i=1}^{5}\sum_{j=1}^{n_i} x_{ij} = \frac{23+27+\ldots+29+30}{40}$$
$$= \frac{1200}{40}$$
$$= 30.$$

Furthermore, the mean of each treatment can be denoted by $\bar{x}_i$; that is $\bar{x}_1 = 28.75$, $\bar{x}_2 = 30.5$, . . ., $\bar{x}_5 = 29.5$.

It could be shown mathematically that the total variation of the data is equal to the sum of variation between treatment groups and variation within treatment groups. In other words:

$$x_{ij} - \bar{x} = (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)$$

The LHS represents the total deviation; the first term in the RHS represents the deviation of treatment mean from the overall mean and the second term in the RHS represents the deviation around treatment mean. If we square both sides this equation, we have:

$$\sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i n_i (\bar{x}_i - \bar{x})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

The first term (LHS) is called total sum of squares, the second term is called sum of squares due to differences between treatments and the third term is called sum of squares due to errors (within treatment). We denote the three terms by the following abbreviations:

$$SSTO = SSTR + SSE$$

SSE is a measure of the random variation of the observations around the respective treatment means. The less variation within treatment, the smaller SSE. If SSE = 0, all observations for a treatment are the same. On the other hand, SSTR measures the extent of differences between treatments, based on the deviations of the treatment means $\bar{x}_i$ around the overall mean $\bar{x}$.

2.2. COMPUTATIONS

Obviously total variation can be calculated as $\sum_j \sum_i x_{ij}^2$, however, we notice that this will be a very large number for a large number of observations. We could subtract this by a correction factor (CF) which is defined as $\frac{1}{N}(grandtotal)^2$, i.e.

$$\begin{aligned} \text{C.F} &= \frac{1}{N}\left(\sum_i \sum_j x_{ij}\right)^2 \\ &= \frac{(1200)^2}{40} \\ &= 36{,}000. \end{aligned}$$

Then total sum of squares is:

$$\begin{aligned} SSTO &= \sum_i \sum_j x_{ij}^2 - CF \qquad [1] \\ &= \left(23^2 + 27^2 + \ldots + 29^2 + 30^2\right) - 36000 \\ &= 696. \end{aligned}$$

The sum of squares due to differences between treatments is:

$$\begin{aligned} SSTR &= \sum_i \frac{\bar{x}_i^2}{n_i} - CF \qquad [2] \\ &= \frac{207^2}{8} + \frac{244^2}{8} + \frac{273^2}{8} + \frac{240^2}{8} + \frac{236^2}{8} - 36000 \end{aligned}$$

or equivalently: $= 8(28.75 - 30)^2 + 8(30.5 - 30)^2 + \ldots + 8(29.5 - 30)^2$

= 276.25.

The sum of squares due to differences within treatment is:

$$\text{SSE} = \sum_i \sum_j x_{ij}^2 - \sum_i \frac{\bar{x}_i^2}{n_i} \qquad [3]$$

= SSTO - SSTR

= 696 - 276.25

= 419.75.

2.3. DEGREES OF FREEDOM

Corresponding to the decomposition of the total sum of squares, we can also obtain a breakdown of the associated degrees of freedom (df). But what is "degree of freedom" ? Well, a rather strict interpretation is that the number of df associated with a chi square variable is the number of independent (standard normal) random variables that conceptually go into the make-up of the variable. For a more intuitive understanding of the term, let us compare two ways of estimating the variance of a population by taking a sample of size *n*: first when we know the value of the population mean μ, and second when we do not know μ.

In the first instance, we estimate the variance by $\sum_{i=1}^{n}(x_i - \mu)^2 / n$; here, the *n* terms $x_i - \mu$ are all independent, hence each makes an independent contribution to the estimation of the variance. Thus we do not lose any degrees of freedom in estimating the variance.

In the second instance, we do not μ, we must replace it by the sample mean $\bar{x}$ and estimate the variance by $\sum_{i=1}^{n}(x_i - \bar{x})^2 / n$. Now recall that $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. This means that the *n* terms $x_i - \bar{x}$ are not independent, because as soon as we know *n*-1 of the terms, the value of the remaining term is fixed. This fact, resulting from our use of an estimate of μ (which is $\bar{x}$) rather than μ itself, causes us to lose one degree of freedom in estimating the variance. Ultimately, we will see that, *in the general problem of estimation, we lose one df for each parameter that is replaced by a sample estimate*.

Now returning to our case:

(a) For SSTO, the calculation was based on 40 observations, but there is one constraint on the deviation $\sum\sum x_{ij} - \bar{x} = 0$, hence it is associated with N-1 = 39 df .

(b) For SSTR, there are 5 treatment groups, but there is one constraint $\sum_i n_i(\bar{x}_i - \bar{x}) = 0$, hence it has 4 def.

(c) For SSE, we can see that expression $\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ which is equivalent to a total sum of squares considering only the *i*th treatment factor. Hence, there are $n_i - 1$ df associated with this sum of squares. So, the number of df for this term in our example is (8-1) + (8-1) + (8-1) + (8-1) + (8-1) = 40 - 5 = 35 df. The SSE is **very important** in the analysis of multiple groups. You can think of it as an average of variances of all treatment groups. Hence a comparison of between groups must be done in relation to this SSE, which we will touch to this statistic in the next discussion.

2.4. SET-UP AN ANOVA TABLE

We can summarise the above computation in an analysis of variance table, commonly known as ANOVA table, as follows:

Source	DF	Sum of squares	Mean of square	F-test
Between treatments	4	276.25	69.06	5.76
Within treatment (residuals)	35	419.75	12.00	
Total	39	696.00		

2.5. THE F TEST

We still have not addressed the question of whether there was any differences between 5 means. The statistical solution to this question is called the F test, named after the eminent British statistician Ronald A. Fisher. The statistic was already defined in Topic 5, which stated that: *if U and V are independently distributed chi-square variables with m and n degrees of freedom (df), respectively, then the ratio* $W = \frac{U/m}{V/n}$ *is distributed according to the F distribution with m and n df.*

Mathematically, it is:

$$\frac{\frac{\sum_{i=1}^{n_1}\left(X_{1j}-\bar{X}_1\right)^2}{\sigma_1^2\left(n_1-1\right)}}{\frac{\sum_{i=1}^{n_2}\left(X_{2j}-\bar{X}_2\right)^2}{\sigma_2^2\left(n_2-1\right)}} = \frac{\hat{\sigma}_1^2/\sigma_1^2}{\hat{\sigma}_2^2/\sigma_2^2} \qquad [4]$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the unbiased estimates of the population variances for population 1 and 2, respectively. Thus, [6] is a function of σ_1^2 and σ_2^2 (the unknown variances). The distribution however holds regardless of the true values of σ_1^2 and σ_2^2. Therefore, under the unique condition (and only such condition) that $\sigma_1^2 = \sigma_2^2$, [5] can be written as:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \qquad [5]$$

Result [5] is often used to test for the equality of two variances.

Now returning to our case. If there is no difference between 5 groups, then one would expected that the **between-treatment** mean square to be small relative to **within-treatment** mean square. On the other hand, if there is difference between treatment groups, the between-treatment mean square should be greater than the within-treatment mean square. We used the word "relative" to talk about differences here. In other words, instead of subtracting one mean square from another, we take the ratio of mean square. And, of course, the ratio of mean square has an F distribution which conveniently allows us to make inference whether a ratio is significantly different from 1 (no difference).

In our example, the mean square due to treatment differences is 69.06 (with 6 df) and the mean square due to within treatment differences is 12.00 (df = 35). Hence the ratio which we denoted as F is:

$$F = \frac{69.06}{12} = 5.76, \qquad \text{with df} = 6, 35.$$

This value is traditionally presented in the last column of the ANOVA Table. Is the ratio significantly different from 1 ? To answer this question, we need to compare this value with the expected value in the F distribution in the appendix. We see that with 6 df in the numerator and 35 (we take 50) df in the denominator and with p=0.01 (1% significance level), the expected F ratio is 3.19. Now, the observed ratio of 5.76 is much larger than this expected ratio, we conclude that there was significant difference between groups.

2.5 ONE-WAY ANALYSIS OF VARIANCE FROM SUMMARY DATA

The above analysis is based on the assumption that individual data are available for each subject. However, suppose that only summarised data are available in the following format:

Group	Sample size	Mean	Variance
1	n_1	$\bar{x}_1$	s_1^2
2	n_2	$\bar{x}_2$	s_2^2
.			
.			
.			
g	n_g	$\bar{x}_g$	s_g^2
Total	N	$\bar{X}$	S^2

First, between-group sum of squares: SSTR = $\sum_{i=1}^{g} n_i(\bar{x}_i - \bar{X})^2$, with (g-1) df

second, the within-group sum of squares: SSE = $\sum_{i=1}^{g}(n_i - 1)s_i^2$, with (N-g) df.

then the ANOVA table can be set up as follows:

Source	DF	Sum of squares
Between groups	g - 1	$\sum_{i=1}^{g} n_i(\bar{x}_i - \bar{X})^2$
Within treatment (residuals)	N - g	$\sum_{i=1}^{g}(n_i - 1)s_i^2$
Total	N - 1	$(N-1)S^2$

III. ANALYSES AND HYPOTHESIS TESTING

Remember we conclude that there is at least one difference between treatment groups in term of weight gains. There are 5 groups, the number of simple two-group comparisons is $C_2^5 = 10$ (in fact, there are many more possible comparisons, can you think of?), then the question is which group is different to which group ? The procedure of searching for pairwise difference is called **multiple comparisons.** There are several procedures for multiple comparisons and they do not necessarily yield the same answer, the crucial issue is we must carefully evaluate them in terms of our aims.

3.1 LINEAR CONTRASTS

This finding of significance signifies the beginning of a careful statistical examination of the results, not the end of the analysis. The search for specific treatment differences involves an application of the method of *multiple comparisons.* The differences between $\bar{x}_1 - \bar{x}_2$, $\bar{x}_2 - \bar{x}_4$, $\frac{\bar{x}_1 + \bar{x}_2}{2} - \bar{x}_4$, $\frac{\bar{x}_1 + \bar{x}_2}{2} - \frac{\bar{x}_3 + \bar{x}_4}{2}$, etc. are only four of literally infinite many comparisons possible among the treatment means.

Each of the comparison can be expressed as a general *contrast*

$$C = \sum_i c_i \bar{x}_i$$

where $c_1, c_2, \ldots, c_n$ are numerical constants so that $\sum_i c_i = 0$.

Thus, for a comparison between $\bar{x}_1 - \bar{x}_2$, we could have $c_1 = 1$, $c_2 = -1$, $c_3 = c_4 = c_5 = 0$. In $\bar{x}_2 - \bar{x}_4$, we could have $c_2 = 1$, $c_4 = -1$ and $c_1 = c_3 = c_5 = 0$. For the comparison $\frac{\bar{x}_1 + \bar{x}_2}{2} - \bar{x}_4$ we can set $c_1 = c_2 = \frac{1}{2}$, $c_4 = -1$ and $c_3 = c_5 = 0$, and so on.

It can be shown that the standard error of a general contrast is:

$$SE(C) = \sqrt{WMS \times \sum_i \frac{c_i^2}{n_i}}$$

where WMS is the within mean square error (in our example WMS = 12.0).

It follows that the ratio

$$L = \frac{C}{se(C)}$$

is distributed according to the *t* distribution with *N*-*g* degrees of freedom (*N* total observation i.e 40 and *g* number of treatments i.e. 5)

3.2 SCHEFFE'S METHOD

Scheffe (1953) is a standard method for multiple comparisons. In this method, a typical contrast C is judged to be statistically significant different from 0 if the absolute value of its associated ratio *L* exceeds *S*, say, where:

$$S = \sqrt{(g-1)F_{g-1,N-g,\alpha}}$$

That is if $|L| < S$, the contrast is not statistically significant.

3.3 TUKEY'S METHOD

If the investigator's interest resides exclusively in pairwise differences between the means, and not in more general comparisons, a criterion proposed by Tukey (1981) is preferable to Scheffe's with respect to power and to the lengths of confidence intervals. It requires, in theory, equal sample sizes (say, $n_1 = n_2 = \ldots = n_5 = n$) and is based on the distribution of the so-called *studentised range*, say:

$$q_{g,N-g} = \frac{\max(\bar{x}_i) - \min(\bar{x}_i)}{\sqrt{\frac{WMS}{n}}}$$

According to Tukey's criterion, the difference between the means of treatment *i* and *j* is significant if:

$$Q_{ij} = \frac{\left|\bar{x}_i - \bar{x}_j\right|}{\sqrt{\frac{WMS}{n}}} > q_{g,N-g,\alpha}$$

and the 95% CI is:

$$\left(\bar{x}_i - \bar{x}_j\right) \pm q_{g,N-g,\alpha}\sqrt{\frac{WMS}{n}}$$

where $q_{g,N-g,\alpha}$ is the tabulated upper α point of the distribution of the studentised range for *g* groups and estimated variance based on *v* df and *n* is the average number of observations per treatment groups.

Example 1 (continued):

For the data in Example 1, we have: $q(5, 35, 0.05) = 4.07$.

and the quantity: $\sqrt{\frac{WMS}{n}} = \sqrt{\frac{12}{8}} = 1.22$.

Hence the critical value for comparisons is: 4.07 x 1.22 = 4.96.

The five sample means are rearranged in ascending order as follows:

$\bar{x}_1 = 25.87$ $\bar{x}_5 = 29.5$ $\bar{x}_4 = 30$ $\bar{x}_2 = 30.5$ $\bar{x}_3 = 34.13$

Obviously: $\bar{x}_3 - \bar{x}_1 = 8.26 > 4.96$; conclusion: significant.

$\bar{x}_3 - \bar{x}_5 = 4.63 < 4.96$; conclusion: no significant; stop.

$\bar{x}_2 - \bar{x}_1 = 4.63 < 4.96$; conclusion: no significant; stop.

and so on. //

3.4 STUDENT-NEWMAN-KEULS (SNK) METHOD

The SNK method provides a modification of the Tukey's method. The test was developed by Newman in 1939 and was generated by Keuls in 1952. Operationally, although the SNK method also makes use of the Studentised range statistic, different critical values are used depending on the number of steps separating the means being tested. To illustrate the difference between the two methods, let us consider the data in Example 1, in which the value of $q(5, 35, 0.05) = 4.07$ is fixed for any comparison. However, for the SNK test, the value of q is dependent on the "distance" between means which are arranged in ascending order.

The test is defined by:
$$q = \frac{\max(\bar{x}_i) - \min(\bar{x}_i)}{\sqrt{\frac{WMS}{n}}}$$

This value is to compared with a critical value of $q(r, N\text{-}g)$ where r is the "distance" between the maximum and minimum means; N is total number of observations and g is the number of treatment groups, i.e. $N\text{-}g$ is the df of the WMS term.

Example 1 (continued):

The means of five treatments are rearranged into ascending order as follows:

Mean: $\bar{x}_1 = 25.87$ $\bar{x}_5 = 29.5$ $\bar{x}_4 = 30$ $\bar{x}_2 = 30.5$ $\bar{x}_3 = 34.13$

Since the distance between $\bar{x}_3$ and $\bar{x}_1$ is 5 steps, therefore, $r = 5$; the distance between $\bar{x}_2$ and $\bar{x}_1$ is 4 steps, therefore $r = 4$, and so on. Based on this procedure we could derive the critical value for any pairwise comparison as follows:

	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$q(r, 35)$	2.89	3.49	3.85	4.10
$W = q(r, 30) \times \sqrt{\frac{12}{8}}$	3.54	4.27	4.71	5.02

Then we can set up the following comparisons:

$\bar{x}_3 - \bar{x}_1 = 8.26 > 5.02$ conclusion: significant; proceed;
$\bar{x}_3 - \bar{x}_5 = 4.63 < 4.71$ conclusion: not significant; stop.

$\bar{x}_2 - \bar{x}_1 = 4.63 < 4.71$ conclusion: not significant; stop.
$\bar{x}_4 - \bar{x}_1 = 4.13 < 4.27$ conclusion: not significant; stop. //

3.5 DUNCAN'S MULTIPLE RANGE TEST

Duncan (1955) developed a procedure for obtaining all pairwise comparisons among *g* sample means. Although this procedure makes use of the Studentised range, his error rate is neither on an experimentwise basis (as with Tukey's) nor on a per-comparisons basis. When the sample means have been ranked from lowest to highest, the error rate is designed in the following way. In general, if two sample means are *r* steps apart, Duncan defines the protection level as:

$$(1-\alpha)^{r-1}$$

the probability of falsely rejecting the equality of two population means when the sample means are *r* steps apart is then taken to be:

$$1-(1-\alpha)^{r-1}$$

For $\alpha = 0.05$, the protection level can be tabulated for various value of *r* as follows:

	Protection level $(1-\alpha)^{r-1}$	Prob of Falsely rejecting Ho. $1-(1-\alpha)^{r-1}$
$r = 2$	0.950	0.050
$r = 3$	0.903	0.097
$r = 4$	0.857	0.143
$r = 5$	0.815	0.185
$r = 6$	0.774	0.226
$r = 7$	0.735	0.265

Because the protection level decreases with increasing r , Duncan's multiple range test is very powerful; that is, there is a high probability of declaring difference when there is actually a difference between population means. This has been one of the reasons for Duncan's test being the most popular among researchers.

According to Duncan, two population means are significantly different if the absolute value of their sample differences exceeds

$$W = q(r, N-g) \times \sqrt{\frac{WMS}{n}}$$

where, as before, n is the number of observations per treatment groups; N is total number of observations from g treatment groups; WMS is the within (residual) mean square derived from the ANOVA table.

Example 1 (continued):

The means of five treatments are rearranged into ascending order as follows:

Mean: $\bar{x}_1 = 25.87$ $\bar{x}_5 = 29.5$ $\bar{x}_4 = 30$ $\bar{x}_2 = 30.5$ $\bar{x}_3 = 34.13$

and the value of $q(r, N\text{-}g)$ are taken from Table 11 then W can be tabulated as follows::

	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$q(r, 35)$	2.89	3.04	3.12	3.20
$W = q(r, 30) \times \sqrt{\frac{12}{8}}$	3.54	3.72	3.82	3.92

Then we can set up the following comparisons:

$\bar{x}_3 - \bar{x}_1 = 8.26 > 3.92$ conclusion: significant; proceed;
$\bar{x}_3 - \bar{x}_5 = 4.63 > 3.82$ conclusion: significant; proceed;
$\bar{x}_3 - \bar{x}_4 = 4.13 > 3.72$ conclusion: significant; proceed;
$\bar{x}_3 - \bar{x}_2 = 3.63 > 3.54$ conclusion: significant; proceed;

$\bar{x}_2 - \bar{x}_1 = 4.63 > 3.82$ conclusion: significant; proceed;
$\bar{x}_2 - \bar{x}_5 = 1.00 < 3.72$ conclusion: not significant; stop.

$\bar{x}_4 - \bar{x}_1 = 4.13 > 3.72$ conclusion: significant; proceed;
$\bar{x}_4 - \bar{x}_5 = 0.50 < 3.54$ conclusion: not significant; stop. //

3.6 FISHER'S LEAST SIGNIFICANCE TEST

The α-level of Fisher's least significance difference (LSD) is valid for a given comparison only if the LSD is used for independent (orthogonal) comparisons or for preplanned comparisons. However, since many people find Fisher's LSD easy to apply and hence use it for making pairwise comparison (particularly those that look "interesting") following the completion of an experiment, the test is recommended only when the overall F-test is significant.

The test is a generalised version of the pairwise t-test, where LSD is defined as:

$$\text{LSD} = \frac{\left|\bar{x}_i - \bar{x}_j\right|}{\sqrt{WMS\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

and is compared with the *t* distribution with *N*-*g* df.

Example 1 (continued):

For the data in Example 1, we have *WMS* = 12 and each of the sample size per treatment are equal to 8, hence:

$$\sqrt{WMS\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = \sqrt{12 \times \frac{2}{8}} = 1.73$$

and with df = 35, we have the critical value for $t(35, 0.05/2) = 2.042$ (notice that we set the a level to be 0.05/2 = 0.025 to have a two-sided hypothesis)

Hence any pairwise difference of more than $(1.73 \times 2.042) = 3.53$ is declared to be significant at the 5% level (2 tailed). For this criteria, we can see that $\bar{x}_3$ and $\bar{x}_1$; $\bar{x}_3$ and $\bar{x}_5$; $\bar{x}_3$ and $\bar{x}_4$; $\bar{x}_3$ and $\bar{x}_2$; $\bar{x}_2$ and $\bar{x}_1$; and $\bar{x}_4$ and $\bar{x}_1$ are significant, whereas there is no statistically significant difference between $\bar{x}_2$ and $\bar{x}_5$; $\bar{x}_4$ and $\bar{x}_5$. //

3.7 SOME COMMENTS

Certainly, a logical question at this point is which one of these tests should one use? Unfortunately, there is no clear-cut answer to this question, and statisticians (like anyone else) often disagree over the utility of the various procedures. Carmer and Swanson (1973) have conducted Monte Carlo simulation studies of a number of multiple comparison procedures, including others not discussed here. They found that the least significance test is a very effective test for detecting true differences in means if applied *only after* the F test in the ANOVA is significant at 5% level. They also report a good performance in detecting true differences with Duncan's multiple range test. This is not surprising, since these two methods are the most powerful of those we have discussed. Duncan's multiple range test is

also available in many computer statistical softwares for ANOVA. It should be satisfactory for many general applications.

Because all multiple comparisons are based on the magnitude of difference two means, we can get some feel for how conservative one test is relative to another by comparing the magnitudes of the differences required for significance to be declared. As can be seen from the above example, Scheffe's procedure is very conservative and should not be used for pairwise comparisons.

3.8 SEVERAL TREATMENTS VS. A CONTROL

It is sometimes necessary to compare several treatment groups versus a control, but not between treatment groups. Dunnett (1964) derived a method for "multiple comparisons with a control" which can be summarised as follows: Suppose that we have p groups each of whose means is to be compared to the same control mean. Let the subscript 0 represent the control group and let n_i ($i = 0, 1, 2, \ldots, p$) denotes the sample size for group i. Dunnett showed that each group can be compared to the control mean by using the statistic

$$L_i = \frac{\bar{x}_i - \bar{x}_0}{s}\sqrt{\frac{n_0 n_i}{n_0 + n_i}}$$

where s is the square root of the pooled variance across all groups.

$$s^2 = \frac{\sum_{i=0}^{p}(n_i - 1)s_i^2}{\sum_{i=0}^{p}(n_i - 1)}$$

with $\sum_{i=0}^{p}(n_i - 1)$ degrees of freedom.

Example 2: The following statistical summary is adapted from data presented by Dunnett in 1964. A total of 60 cockerels were assigned at random to receive either no treatment (control) or one of the drugs in their diets. These 60 birds were sacrificed at either 1, 3 or 7 weeks after the start of treatment, and the fat content of the breast muscle was measured; the time of sacrifice had no effect on the response.

Group	Sample size	Mean	S.D	Li
0. Control	15	2.580	0.258	
1. Stilbesterol	15	2.461	0.409	-0.92
2. Low dose acetyl enheptin	15	2.232	0.381	-2.70
3. High dose acetyl enheptin	15	2.573	0.348	-0.05

Since the four sample sizes were equal, the overall (pooled) variance can be calculated as : $s^2 = \frac{1}{4}\sum s_i^2 = 0.125$, with 4 x 14 = 56 degrees of freedom. The value of L_i should be referred to the critical value table by Dunnett (1955, 1964) for assessment of significance.

If, however, the sample size for each group is NOT equal to the control group, a conservative upper bound seems to be

$$m \le 1 + 0.07\left(1 - \frac{n_i}{n_0}\right)$$

For example, if $n_0 = 20$ and $n_i = 10$, the correct critical value for judging the significance of L_i should be no more than 1 + 0.07(1 - (10/20)) = 1.035 times the tabulated value.

IV. TWO-WAY ANALYSIS OF VARIANCE

In the above analysis, we were concerned with classification of observation by a single criterion (factor) and our primary purpose was to test for the differences between levels of the factor. We shall now consider the situation in which the individual observations are subject to two criteria of classification. Let us start with a simple illustration. Suppose that the means of a certain variable trait for three groups of subjects and two treatments are as follows:

	Group 1	Group 2	Group 3
Treatment A	28	33	35
Treatment B	28	33	35

For this data, we may denote each value for an *i*th row and *j*th column as x_{ij} (i = 1, 2 and j = 1, 2, 3). The overall mean (denoted by $\overline{X}$) is $\overline{X}$ = 32. We can say that the effects associated with groups 1, 2 and 3 are 28 - 32 = -4; 33 - 32 = 1 and 35 - 32 = 3, respectively. If we denote these effects by α_j (j = 1, 2, 3) then $\alpha_1 = -4$, $\alpha_2 = 1$ and $\alpha_3 = 3$, and note that the columns (groups) of the above table differ from each other, but the rows (treatments) within each column show identical values. In fact, each value can be represented by the equation:

$$x_{ij} = 32 + \alpha_j$$

Consider now another scenario, where there is no effects associated with the patient groups but there are effects associated with treatments. The data may look like:

	Group 1	Group 2	Group 3
Treatment A	34	34	34
Treatment B	30	30	30

So, the effect of treatment A is now 34 - 32 = 2 and of treatment B is 30 - 32 = -2. Each value fits the equation

$$x_{ij} = 32 + \beta_i$$

where $\beta_1 = 2$ and $\beta_2 = -2$.

Now suppose that there are *both* treatment effects and differences between patient groups. We may simulate the data according to the equation

$$x_{ij} = 32 + \alpha_j + \beta_i$$

then the data may look like:

	Group 1	Group 2	Group 3
Treatment A	32 + 2 - 4 = 30	32 + 2 + 1 = 35	32 + 2 + 3 = 37
Treatment B	2 - 2 - 4 = 26	32 - 2 + 1 = 31	32 - 2 + 3 = 33

In this case, the six subgroups yield means differing across the different cells of the table. However, the effect of a combination, $\bar{x}_{ij} - \bar{X}$, associated with cell *ij* is exactly equal to the effect associated with its row, β_i, plus the effect associated with its column, α_j. In this scenario, we say the effect is *additive*.

Now, we will consider the case where the effect is not additive but *interactive* (interaction effect):

	Group 1	Group 2	Group 3
Treatment A	32 + 2 - 4 - 2 = 28	32 + 2 + 1 + 6 = 41	32 + 2 + 3 - 4 = 33
Treatment B	2 - 2 - 4 + 2 = 28	32 - 2 + 1 - 6 = 25	32 - 2 + 3 + 4 = 37

The effect associated with a cell is now no longer the simple sum of the effects of its row and its column, an indication that interaction effects are present. **Notice that columns are different in different ways within rows, and vice versa, when interaction is present**. Naturally, for any real data, there will be random error as well. This implies that the problem is now threefold: we must find out (1) if there are effects of the treatments represented by rows, (2) if there are effects associated with columns and (3) if there are effects which are attributable neither to rows (irrespective of columns) nor columns (irrespective of rows) but rather to interaction.

A general set of observed value for the two-variable classification problem can be formatted as follows:

	Treatment group				Mean
Block	1	2	3 . . .	g	all groups
1	x_{11}	x_{12}	x_{13} . . .	x_{1g}	$\overline{x}_{1.}$
2	x_{21}	x_{22}	x_{23} . . .	x_{2g}	$\overline{x}_{2.}$
3	x_{31}	x_{32}	x_{33} . . .	x_{3g}	$\overline{x}_{3.}$
.					
.					
.					
n	x_{n1}	x_{n2}	x_{n3} . . .	x_{ng}	$\overline{x}_{n.}$
Mean	$\overline{x}_{.1}$	$\overline{x}_{.2}$	$\overline{x}_{.3}$ . . .	$\overline{x}_{.g}$	$\overline{X}$

That is, we have g treatments groups and in each group, we have n subjects. So, each observation can be identified by x_{ij}, where $i = 1, 2, 3, \ldots, g$ and $j = 1, 2, 3, \ldots, n$. In other words, total number of observations is

$$N = ng$$

Obviously the total sum of squares is measured by:

$$\text{SSTO} = \sum_{i=1}^{g}\sum_{j=1}^{n}\left(x_{ij} - \bar{X}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{ng}\left(\sum_{i=1}^{g}\sum_{j=1}^{n} x_{ij}\right)^2$$

which is associated with N-1 df.

It can be shown that this sum squares can be partitioned into three sources, namely, between blocks, between treatments and residual errors, as follows::

Between block:

$$\text{SSB} = g\sum_{i=1}^{n}\left(\bar{x}_{i.} - \bar{X}\right)^2 \qquad \text{with} \quad n\text{-1 df}$$

Between treatment groups:

$$\text{SSTR} = n\sum_{j=1}^{g}\left(\bar{x}_{.j} - \bar{X}\right)^2 \qquad \text{with } g\text{-1 df.}$$

Residual error:

$$\text{SSE} = \sum_{i=1}^{g}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{X}\right)^2$$

$$= \text{SSTO - SSB - SSTR} \quad \text{with } (g\text{-1})(n\text{-1}) \text{ df}$$

These information can be tabulated in an ANOVA table as follows:

Source	DF	Sum of squares
Between groups	g - 1	$n\sum_{j=1}^{g}\left(\bar{x}_{.j} - \bar{X}\right)^2$
Between blocks	n - 1	$g\sum_{i=1}^{n}\left(\bar{x}_{i.} - \bar{X}\right)^2$

Residual errors	$(g\text{-}1)(n\text{-}1)$	$\sum_{i=1}^{g}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{X}\right)^2$
Total	ng - 1	$\sum_{i=1}^{g}\sum_{j=1}^{n}x_{ij}^2-\frac{1}{ng}\left(\sum_{i=1}^{g}\sum_{j=1}^{n}x_{ij}\right)^2$

Example 3: The following table shows how the data from a randomised block study would be arranged for analysis. The notation in the row and column headed means speaks for itself. The standard deviation are presented in the last row of the table for two reasons: one is to bring to the reader's attention the importance of the assumption of equal variances for a fully informative analysis.

Clotting time of plasma (in minutes) for 4 treatments compared in a randomised clinical trial.

	Treatment group				Mean
Subject	1	2	3	4	all groups
1	8.4	9.4	9.8	12.2	9.95
2	12.8	15.2	12.9	14.4	13.825
3	9.6	9.1	11.2	9.8	9.925
4	9.8	8.8	9.9	12.0	10.125
5	8.4	8.2	8.5	8.5	8.40
6	8.6	9.9	9.8	10.9	9.8
7	8.9	9.0	9.2	10.4	9.375
8	7.9	8.1	8.2	10.0	8.55
Mean	9.3	9.7125	9.9375	11.025	9.9938
SD	1.55	2.294	1.514	1.815	

The analysis will begin with a calculation of total sum of squares:

$$\begin{aligned}\text{SSTO} &= (8.4-9.9938)^2+(12.8-9.9938)^2+\ldots+(10-9.9938)^2\\ &= 105.7788\end{aligned}$$

Between treatment SS:

$$\begin{aligned} \text{SSTR} &= 8\left[(9.3-9.9938)^2+\ldots+(11.025-9.9938)^2\right] \\ &= 13.0163 \end{aligned}$$

Between subjects SS:

$$\begin{aligned} \text{SSB} &= 4\left[(9.95-9.9938)^2+\ldots+(8.55-9.9938)^2\right] \\ &= 78.9888 \end{aligned}$$

And residual SS:

$$\begin{aligned} \text{SSE} &= 105.7788 - 13.0163 - 78.9888 \\ &= 13.7737 \end{aligned}$$

And the ANOVA table can be setup fully as follows:

Source	DF	Sum of squares	Mean of square	F-test
Between treatments	3	13.0163	4.3388	6.62
Between subjects	7	78.9888	11.2841	
Residuals	21	13.7737	0.6559	
Total	31	105.7788		

Since the F ratio (6.62) exceeds F(3, 21, 0.05) = 3.07, we conclude that statistically significant differences exist among the treatment means at the 5% level. Multiple comparison may be made by using any of the criteria presented in the earlier section.

Let $c_1, c_2, \ldots, c_g$ denote a set of constants with condition $\sum c_j = 0$, the contrast

$$C = \sum c_j \bar{x}_{.j}$$

can be tested by the Scheffe's criterion as

$$L = \frac{C\sqrt{n}}{\sqrt{WMS \times \sum_{j=1}^{g} c_j^2}}$$

If $L > \sqrt{(g-1)F_{g-1,(g-1)(n-1),\alpha}}$ then the difference would be declared to be significant.

On the other hand, one can use the Bonferroni's criterion which would lead to judgement of significance if

$$|L| > t_{g-1,(n-1)(g-1),\frac{\alpha}{2k}}$$

where k is the number of prescribed comparisons.

All pairwise comparisons (Tukey's, Scheffe's, Duncan's, LSD etc) can be proceeded as described in previous sections.

V. EXERCISES

1. What are the differences between sum of squares, mean square, variance and variation?

2. In a multicentre clinical trial which compared the efficacy of two drugs A and B. The trial was carried in 4 different countries. Patients were classified by sex (male or female) and within each sex patients were stratified into 4 age groups. In other words, there are 32 different comparisons the efficacy of treatments A and B. Assuming that if the p value for each comparison is less than 0.05, we declare a statistical significance. If there was actually no difference between the treatments, what is the probability that one subgroup comparison will reveal a statistical significance. How many comparisons which would be significant by chance alone would we expect from 32 comparisons?

3 Consider the following experiment, which compared a new antiinflammatory drug N with aspirin and placebo. There were 11 subjects in each treatment group; giving a total of 33 subjects. Each subject was a definite rheumatoid arthritis patient. The response measured was an index of treatment effectiveness:

Patient No.	Placebo	Aspirin	N
1	1.0	1.3	2.1
2	-0.6	2.7	1.1
3	0.7	2.1	2.4
4	1.4	0.7	0.1
5	1.0	3.6	0.1
6	1.8	1.9	-0.1
7	0.2	3.9	-0.3
8	1.7	-0.8	0.8
9	0.4	2.2	-0.6
10	1.0	1.9	0.6
11	0.2	2.8	0.3

Mean	0.80	2.03	0.59
SD	0.72	1.32	0.95

Perform an analysis of variance and test the hypothesis of equality of treatment means. Also, compare the multiple comparison procedures as described in part II in the note.

4. The following data represents a randomly selected twin pairs (MZ and DZ) from our data base.
(a) Perform an analysis of variance to decompose the within and between pair variations fro MZ and DZ pairs separately.
(b) Perform an analysis of variance to test whether there was any effect of VDR genotypes (BSM). What would you do in terms of the paired twin data.

Pair	Zygosity	Bsm1	Bsm2	LS1	LS2
12	MZ	AA	AA	1.020	0.960
21	MZ	AA	AA	1.060	1.060
68	MZ	BB	BB	1.061	1.071
98	MZ	AB	AB	1.301	1.337
82	MZ	AA	AA	1.142	1.131
56	MZ	AA	AA	0.790	0.740
16	MZ	AB	AB	1.110	1.060
29	MZ	AB	AB	1.248	1.240
84	MZ	BB	BB	1.150	1.240
106	MZ	AA	AA	1.112	1.138
75	MZ	AB	AB	1.006	1.069
76	MZ	AB	AB	1.150	1.140
18	MZ	AB	AB	0.980	1.020
55	MZ	AA	AA	1.150	1.180
11	MZ	AB	AB	1.120	1.040
100	MZ	AB	AB	1.317	1.336
66	MZ	AB	AB	0.984	1.036
80	MZ	AA	AA	1.147	1.100
1	MZ	AB	AB	1.095	1.010
92	MZ	AA	AA	1.060	0.950
20	MZ	BB	BB	1.300	1.360
89	MZ	AA	AA	1.000	1.074
81	MZ	BB	BB	1.110	1.100
31	MZ	AB	AB	1.080	1.112
13	MZ	BB	BB	1.483	1.080
73	MZ	BB	BB	1.297	1.358
45	MZ	AB	AB	1.420	1.470
36	MZ	BB	BB	1.360	1.350
25	MZ	AB	AB	0.870	0.870
17	MZ	AB	AB	1.210	1.190
8	MZ	AB	AB	1.010	1.000

107	DZ	AB	AB	1.370	1.120
54	DZ	AA	AA	1.370	1.200
46	DZ	AB	AB	1.350	1.330
34	DZ	AA	AA	1.030	1.100
70	DZ	BB	BB	1.240	1.040
50	DZ	AB	AB	1.130	1.250
71	DZ	AB	AB	1.190	1.210
58	DZ	AA	AA	1.080	1.119
42	DZ	BB	BB	1.420	1.390
52	DZ	AB	AB	1.160	1.280
41	DZ	AB	AB	1.110	1.160
132	DZ	AB	AB	1.209	1.254
124	DZ	AA	AA	1.098	1.112
57	DZ	AA	AA	0.830	1.110
26	DZ	BB	BB	1.100	1.140
101	DZ	AB	AB	1.060	1.280
112	DZ	AB	AB	1.237	1.325
110	DZ	BB	BB	1.210	1.213
113	DZ	BB	BB	1.297	1.324
121	DZ	AB	AB	1.272	1.234
97	DZ	BB	BB	1.200	1.312
83	DZ	AB	AB	1.231	1.254
109	DZ	AB	AB	1.331	1.391
64	DZ	AB	AB	1.290	1.240
67	DZ	AB	AB	1.220	1.360
3	DZ	AB	AB	1.140	1.140
51	DZ	AA	AA	1.330	1.300
62	DZ	AB	AB	1.100	1.360
32	DZ	AB	AB	1.050	1.000
44	DZ	BB	BB	1.130	1.140
15	DZ	AB	AB	1.240	1.260
78	DZ	BB	BB	1.213	1.300
86	DZ	BB	BB	1.290	1.040
105	DZ	BB	AB	1.310	1.080

(This data is in the Bone Network called "ANOVA Exercise".

5. A study of the effect of VDR genotypes on bone loss among postmenopausal women in Switzerland (Lancet 1995) reported the following results:

VDR Genotype	Sample size (N)	% Bone loss (Mean $\pm$ SE)	Age (mean $\pm$ SD)
bb	26	0.7 ± 0.7	70.5 ± 6.2
Bb	37	1.0 ± 0.7	73.8 ± 7.0
BB	9	-2.3 ± 1.0	72.7 ± 9.5

The authors wrote: "the rate of chance was significantly greater ($p<0.05$, ANOVA) in homozygote BB (0.7% per year) than in homozygote bb (-2.3% per year) or in heterozygote Bb subjects (1% per year)".

How would you assess the baseline compatibility? Perform an analysis of variance based on summary data and calculate necessary statistics (tests and confidence interval). Do you agree with the authors' conclusion?

6. The computer output shown here gives the ANOVA for a taste-test experiment where each of 12 persons was asked to sample three new formulations of a widely used bulk laxative:

```
MTB> TWOWAY ANOVA OF "RATING" BY "FORM" AND "PERSON"

ANALYSIS OF VARIANCE RATING
       SOURCE    DF      SS       MS
       FORM       2     767.4   383.7
       PERSON    11    5301.2   481.9
       ERROR     22    1532.6    69.7
       TOTAL     35    7601.2

       MTB> TABLE BY "FORM"
       SUBC> MEANS OF "RATING"
       ROW: FORM
          RATING MEAN
       1        57.667
       2        46.917
       3        49.250
       ALL      51.278
```

(a) Describe the factors and sample sizes involved in this experiment.
(b) Calculate an F test to compare the formulation means. Give the p-value for your test.
(c) Calculate 95% CI for all pairwise differences in formulation means. Which formulation means appear to be different?

7. Consider an experiment to compare three treatments A, B and C. It was decided 6 replicates were necessary, so 3 patients were selected from each of 6 age x sex groups, to form 6 blocks of 3. Treatments were then allocated randomly within each block. This is called a **randomised complete block** design. The response is as follows:

			Treatment group			
Age	Sex	Block	A	B	C	Mean
1	M	1	7.9	12.4	13.1	11.1
2	M	2	8.8	14.0	13.8	12.2
3	M	3	13.0	14.6	14.1	13.9
1	F	4	8.8	8.8	13.7	10.4
2	F	5	10.0	12.6	13.9	12.2
3	F	6	12.2	12.0	14.0	12.7
		Mean	10.1	12.4	13.8	

Perform an ANOVA and test whether there is any difference between treatment groups. Give 95% confidence interval of differences.