ANALYSIS OF VARIANCE, ANALYSIS OF COVARIANCE AND MULTIVARIATE ANALYSIS OF VARIANCE

T. Krishnan
Cranes Software International Limited, Mahatma Gandhi Road, Bangalore - 560 001
krishnan.t@systat.com

1. Estimating Variance

Analysis of variance, which is usually abbreviated by ANOVA, is a technique of decomposing the total variation present in data in terms of variation due to identifiable and interesting sources that cause this variation. Let us start with an example. The following figures represent the yields (in bushels per acre) of wheat in 12 farms in a district:

96 93 37 81 58 79 69 101 73 96 81 102

The overall mean is 80.5. To compute the variance we first compute

$$(96-80.5)^2 + (93-80.5)^2 + \ldots + (102-80.5)^2 = 4109$$

and then divide it by 11 to get

$$4109/11 = 373.545,$$

as variance. The numerator is called the SUM OF SQUARES. You must have learnt the reason for dividing by 11 rather than 12. Now this number 80.5 represents our estimate of the mean of yield from a population of farms in this district if these 12 farms form a random sample; similarly 373.545 is our estimate of the variance of yield among farms in this district.

2. Comparison of Two Means

The foregoing exercise is carried out on the basis that the farms in the district are homogeneous in respect of yield and the variation observed is due to causes unknown, unexplainable, due to chance, or just not investigated. Suppose, however, there are two varieties of wheat used in this district and the data are actually as follows:

Variety 1: 96 37 58 69 73 81
Variety 2: 93 81 79 101 96 102

We envisage the possibility that the mean yields of these two varieties are different. Then we consider farms under these two varieties as two different statistical populations. The mean yields are then estimated by 69 and 92 respectively of Variety 1 and Variety 2.

Let us assume (for convenience) that the variances in these two populations are however the same. Remember, you made this assumption when you did the two-sample t-test for comparing two (normal) means. Also, remember that such assumptions should be tested before or after the analysis depending upon the diagnostic methods used. Anyhow, how will you estimate the common variance now? You do what you did in the two-sample t-test to compute the denominator of the t statistic; you take deviations from the respective means, and not from the common mean, since the mean is not common to the two varieties. This will give the SUM OF SQUARES now as

$(96\text{-}69)^2 + (37\text{-}69)^2 + \ldots + (81\text{-}69)^2 + (93\text{-}92)^2 + (81\text{-}92)^2 + \ldots + (102\text{-}92)^2 = 2522.0.$

If we want to estimate the common variance now, we use this and divide by 10. The reason for dividing by 10 is similar to the reason for dividing by 11 earlier. Earlier we estimated a single overall mean and so deducted one from the number of observations to get the denominator for the sum of squares; now we have estimated two means and so we deduct two from the total number of observations to get 10 for the denominator. So now the estimate for the common variance is 2522.0/10 = 252.2.

If the means of the two varieties are different then the latter is a valid estimate of common variance and the former is not. If the means of the two varieties are the same, notice that each of the deviations is the same in the former sum of squares as well as in the latter, and so the two sums of squares will be the same.

We call the numbers 11 and 10 respectively DEGREES OF FREEDOM of the corresponding Sum of Squares. These estimates of variance are of the form:

Sum of Squares/Degrees of Freedom

and are called MEAN SQUARES. Different mean squares give estimates of variance under different conditions or of different variances.

What we have done so far is to have decomposed the
TOTAL SUM OF SQUARES of 4109.0 with 11 DEGREES OF FREEDOM into SUM OF SQUARES 1 of 2522.0 with 10 DEGREES OF FREEDOM and SUM OF SQUARES 2 of 1587.0 with 1 DEGREE OF FREEDOM, the last one computed as the difference between the total sum of squares and Sum of Squares 1. Let us abbreviate Sum of Squares by SS, Mean Squares by MS. and Degrees of Freedom by DF.

What is the meaning of this decomposition? Let us look at two observations, say 96 from Variety 1 and 101 from Variety 2. Why are they different? One identifiable reason is that the varieties are different. Okay, is that the only reason or source of variation? If so, how can you explain the two observations 96 and 37 from the same variety being different? Evidently, because they are different farms, where the conditions like fertilizer, soil, water, weather, etc. are different, even if the variety is the same. But these other factors were not controlled in the experiment (not always a good way to design an experiment) and so we attribute these differences to chance. Thus the difference between 96 and 101 from the two varieties has two sources of variation---one due to variety and another due to chance (variously called natural, experimental, error or residual variation). Thus since in SS1 you took deviation from the corresponding variety mean, it is a measure of the chance or RESIDUAL source of variation.

What is the other source of variation? From the discussion it is evident that it is due to the difference in the means of the two varieties. How do you compute the SS due to it? Let us look at the two means. They are 69 and 92 and the mean of the means is 80.5 and the variance of each mean $\frac{1}{6}$ of the common variance. Thus

$$(69\text{-}80.5)^2 + (92\text{-}80.5)^2 = 264.5$$

(divided by 1) is an estimate of $\frac{1}{6}$ of the common variance and this estimates the common variance as 6 times 264.5 = 1587.0. This exactly is the SS2, which can be attributed to Varietal difference with 1 DF. Thus we have decomposed the total sum of squares into two components, one due to Variety and another due to residual or natural variation. Then we ask the question: Is the variation due to variety of the same magnitude as that due to the residual? This asks the question if the difference in the observations under Variety 1 and Variety 2 is of the same magnitude as that between two observations under the same Variety? This can now be examined by looking at the ratio of the Mean Squares due to Variety and Residual, investigating its sampling distribution under the hypothesis of no difference in means, and using it calculate the probability (p-value) corresponding to the observed ratio. This is precisely what an ANALYSIS OF VARIANCE does. The Analysis of Variance is usually presented in a table of the following form:

Source	SS	DF	MS	F-ratio	p-value
Variety	1587	1	1587	6.292	0.031
Residual	2522	10	252.2		

From the foregoing discussion, you can write:
Observation i on variety 1, y_{1i} = mean of variety 1 + residual variation e_{1i}
Observation i on variety 2, y_{2i} = mean of variety 2 + residual variation e_{2i}

The residual variations are all assumed to be statistically independent with mean 0 and a common variance (denoted by σ^2) and when required, also assumed to be normally distributed. All these statements together form a MODEL.

Now with these computations, we can examine the difference between the two varieties. The difference Variety 1 – Variety 2 mean is 92-69 = 23. What is the variance of this estimate? It is

$$252.2\left(\frac{1}{6} + \frac{1}{6}\right) = 84.0.$$

And the standard error of this difference is $\sqrt{84.0}$ = 9.1688, leading to the t-value of 2.5085 with 10 DF with a p-value of 0.031 and 95% confidence interval of (2.57, 43.43) and 99% confidence interval of (-6.06, 52.06). Thus we reject the hypothesis of equal means at 5% (since 0 is not in the interval) and do not reject at 1% (since 0 is in the interval). This can also be concluded from the p-value of 0.031, by which we reject the hypothesis of equality of means at level 0.05 (5%) and do not reject at level 0.01 (1%). Incidentally, the F-ratio 6.292 in the table is the ratio of the Variety MS and the Residual MS, has a numerator DF and a denominator DF; in this case the numerator DF is 1 and the denominator DF is 10. The F(1, k) statistics is the square of the t-statistic with k DF. Here the F-statistic 6.292 is the square of the t-statistic of 2.5085. So the analysis of variance test in this case of two means, is the same as the t-test.

3. Comparison of More Than Two Means

Now what if we have more than two varieties? Consider the following data on four varieties from Sahai and Ageel (2000):

Variety1	Variety 2	Variety 3	Variety 4
96	93	60	76
37	81	54	89
58	79	78	88
69	101	56	84
73	96	61	75
81	102	69	68

The four means are:

69	92	63	80

We make the same assumptions as before, that is, the model is the same except that now there are four varieties and so four varietal means, estimated by the means above. The common residual variance is estimated as before as follows:

Residual Sum of Squares:

$(96-69)^2 + (37-69)^2 + \dots + (81-69)^2 + (93-92)^2 + (81-92)^2 + \dots + (102-92)^2 + (60-63)^2 + (54-63)^2 + \dots + (69-63)^2 + (76-80)^2 + (89-80)^2 + \dots + (68-80)^2 = 3272.0$

which when divided by DF 20 (calculated by the same argument as 24-4 = 20) gives

$$3272.0/20 = 163.6$$

as an estimate of the common variance σ^2. This estimate is different from the earlier estimate of the same σ^2 since it is now based on 24 observations.

The variety SS is computed in the same manner as before but from four means with three DF. The overall mean now is 76. Thus the variety SS is

$$6[(69-76)^2 + (92-76)^2 + (63-76)^2 + (80-76)^2] = 6 \times 490 = 2940$$

with 3 DF since four means are compared using their overall mean. Thus the Variety MS = Variety SS/3 = 2940/3 = 980. Using this the F(3, 20)-ratio is 980/163.6 = 5.99 with a p-value of 0.00439 showing a great deal of significance even at 1%. The Analysis of Variance table is presented as follows:

Source	SS	DF	MS	F-ratio	p-value
Variety	2940	3	980	5.990	0.004
Residual	3272	20	163.6		

This analysis cannot be interpreted in terms of the t-test since four means are compared here whereas t-test compares two means like the first F-test. F-statistic with a numerator DF > 1 does not have a t-statistic analogue.

4. Other Comparisons

This test shows that the four varieties have different average yields. Now one may be interested in finding out where these differences lie. Several questions may be asked in this context: Is Variety 1 different from Variety 3? Is Variety 2 different from Variety 4? Is the average of Varieties 1 and 3 different from the average of Varieties 2 and 4? Is Variety 1 different from average of Varieties 2, 3, and 4? And so on. All these questions can be answered using the Residual MS computed in the ANOVA table above. Although you can compare Variety 1 and 3 by a t-test or an ANOVA test based on the 12 observations (with 10 DF for t) on Varieties 1 and 3, it is better to use the 20 DF denominator from the latter ANOVA table (with 20 DF for t) since this is based on all the observations, and hence is a better estimate of σ^2 (under the assumption of a common σ^2).

Thus SS for Variety 1 vs Variety 3 comparison is:

$$t^2 = 3(69-63)^2/163.6 = 108/163.6 = 0.6601, \text{ p-value} = 0.4261$$

and hence the 1 DF SS due to this comparison is 108.

Similarly Variety 2 vs Variety 4 comparison is:

$$t^2 = 3(92-80)^2/163.6 = 432/163.6 = 2.6406, \text{ p-value} = 0.1198$$

and hence the 1DF SS due to this comparison is 432.

How about comparing 1 & 3 on the one hand and 2 & 4 on the other?

Here you compute (69+63)/2 and (92+80)/2 and compare. Since (69+63)/2 and (92+80)/2 are each based on 12 observations the variance of the difference (69+63)/2 - (92+80)/2 is $\sigma^2/6$. Thus for this comparison

$$t^2 = 6(66-86)^2/163.6 = 2400/163.6 = 14.66, \text{ p-value} = 0.0010.$$

Thus for the observed differences among the four varieties the difference between 1 & 3 vs 2 & 4 is a main contributor, 2 vs 4 some, and 1 vs 3 a tiny bit. Notice that the three 1 DF SS 108, 432, 2400 add up to the 3 DF Variety SS 2940 in the ANOVA table. This is because these are statistically ‘independent’ comparisons.

With these computations, the ANOVA table can be extended in the following form:

Source	SS	DF	MS	F-ratio	p-value
Variety 1 vs 3	108	1	108	0.6601	0.4261
Variety 2 vs 4	432	1	432	2.6406	0.1198
Variety 1,3 vs 2,4	2400	1	2400	14.66	0.0010
Variety	2940	3	980	5.990	0.004
Residual	3272	20	163.6		

Note that unless the components are ‘statistically’ independent, the SS do not add up like in the table above. It is not always that we are interested in investigating independent comparisons. For instance we may be interested in the comparisons 1 vs 2, 1 vs 3, 2 vs 3, which are not 'statistically' independent.

Suppose 1 is a control and 2, 3, and 4 are experimental varieties and you want to compare control 1 (with mean 69, variance $\sigma^2/6$) vs 2, 3 & 4 (with mean 78.33, variance $\sigma^2/18$). Then the difference has variance σ^2 (1/6 + 1/18) = (2/9)σ^2 and hence

$$t^2 = (9/2)\,(69-78.33)^2/163.6 = 312.25/163.6 = 1.9086 = 0.1824.$$

All these t-statistics have 20 DF and the t^2-statistics are F(1, 20).

5. Contrasts

The examples we have given above of comparisons are called Contrasts. A Contrast is a linear combination of means μ_i with coefficients α_i:

$$\alpha_1\mu_1 + \alpha_2\mu_2 + \ldots + \alpha_k\mu_k$$

where $\alpha_1 + \alpha_2 + \ldots + \alpha_k = 0$. The above hypotheses we tested have the form:

$$H_0: \alpha_1\mu_1 + \alpha_2\mu_2 + \ldots + \alpha_k\mu_k = 0$$

The test statistic for a contrast is similar to that for a two-sample *t* test; the result of the contrast (a relation among means, such as mean *A* minus mean *B*) is in the numerator of the test statistic, and an estimate of within-group variability (the pooled variance estimate or the error term from the ANOVA) is part of the denominator.

You can select contrast coefficients to test:

- Pairwise comparisons (test for a difference between two particular means)
- A linear combination of means that are meaningful to the study at hand (compare two treatments versus a control mean)
- Linear, quadratic, or the like increases (decreases) across a set of ordered means (that is, you might test a linear increase in yield due to fertilizer levels used in a limited range).

6. Multiple Comparisons

Suppose you are interested in making several comparisons at a time in your quest to find out where the differences lie and/or to order the varieties in decreasing order of average yields. There is a problem here. Remember that if the hypothesis of no differences in the means holds, and if you are testing at 5% level of significance the chance that you will (wrongly) reject the hypothesis is 0.05. Suppose you test all pairs of the four varieties. There are six comparisons: 1 vs 2, 1 vs 3, 1 vs 4, 2 vs 3, 2 vs 4, 3 vs 4. If you test each pair at 5% level, then even if the hypothesis of no differences holds, the chance of rejecting one or more of these 6 comparisons is much larger than 5%. An approximate calculation (assuming independence of the tests, which is not quite correct) is:

$$1-(1-0.05)^6 = 1-(0.95)^6 = 1-0.7351 = 0.2649$$

much higher than 0.05. Thus there is a fair chance that you will find significance of some pair even when all the four means are the same. What you are doing while simultaneously testing several hypotheses is to increase the level of significance of the overall test---in this case of 6 tests at 5% to something like 26%. This is the problem of multiple comparisons. The solution to this problem is not easy and a host of procedures exist suitable for different situations. Most of them attempt to decrease the level of significance of each test so that the overall level is at the intended level. For instance if you decrease the level of each test to 1%, the above computation will give the overall level of approximately 5.85%. These procedures are also called *post hoc* tests.

The results of all 6 pairwise comparisons are as follows:

Variety i	vs Variety j	Mean (i-j)	Std Error	Significance
1	2	23	7.385	0.032
	3	- 6	7.385	1.000
	4	11	7.385	0.912
2	3	29	7.385	0.005
	4	12	7.385	0.719
3	4	17	7.385	0.193

The results in an ANOVA table serve only to indicate whether the means differ significantly or not. They do not indicate which mean differs from which other. To report the pairs of means that differ significantly, you might think of computing a two-sample *t* test for each pair; however, do *not* do this, for reasons explained above. The result of following such a strategy is to declare differences as significant when they are not.

As an alternative to the situation described above, several techniques are available to perform pairwise mean comparisons controlling individual test's error rates or the error rate of the entire family of tests. You have to choose a proper test based on the variance assumptions and the type of error rate to be controlled. The following tests divided into two parts based on variance assumptions, are some of the standard tests:

Equal Variance	Unequal Variance
Tukey	Tamhane's T2
Bonferroni	Games-Howell
Fisher's LSD	Dunnett's T3
Sidak	
Scheffé	
Tukey's b	
Duncan	
Ryan-Einot-Gabriel-Welsch Q	
Hochberg's GT2	
Gabriel	
Student-Newman-Keuls	
Dunnett	

The Student-Newman-Keuls Procedure (S-N-K) and Duncan's multiple range test control neither the individual nor the family-wise error rates. Duncan's test has been heavily criticized in the statistical literature; it gives many more statistically significant differences than is warranted and does not really protect the significance level. As a general rule, Fisher's LSD is one of the more liberal procedures (more likely to declare means different), but it does not control the family-wise error rate. Tukey's and Scheffé's methods are conservative, with Scheffé's method being more conservative than Tukey's method.

There is an abundance of literature covering multiple comparisons (see Miller, 1985); however, a few points are worth noting here:

- If you have a small number of groups, the Bonferroni pairwise procedure will often be more powerful (sensitive). For more groups, consider the Tukey method. Try all the methods in ANOVA (except Fisher's LSD) and pick the best one.
- Carrying out all possible pairwise comparisons is a waste of power. Think about a meaningful subset of comparisons and test this subset with Bonferroni levels. To do this, divide your critical level, say 0.05, by the number of comparisons you are making. You will almost always have more power than with any other pairwise multiple comparison procedure.
- Some popular multiple comparison procedures do not maintain their claimed protection levels. Other stepwise multiple range tests, such as the Student-Newman-Keuls and Duncan's tests, have not been conclusively demonstrated to maintain overall protection levels for all possible distributions of means.
- Some tests produce and test homogeneous subsets of group means instead of testing each pair of the group means.
- Some tests come under unequal variance assumptions and use group variances instead of MSE to compare the group means.

7. Two-Factor Analysis of Variance

Suppose you have more than one known or designed factor or source of variation in your data; say, in this example, two different fertilizers are used and the data are as follows:

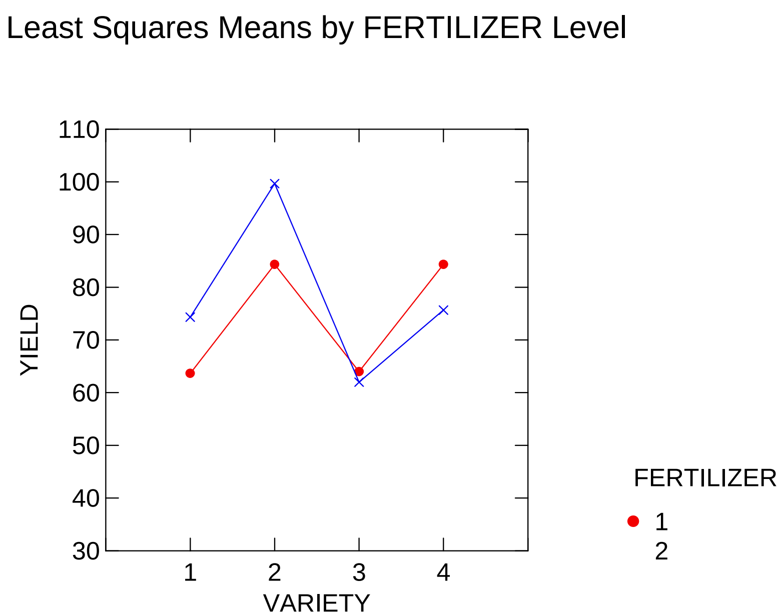
	Variety 1	Variety 2	Variety 3	Variety 4
Fertilizer 1	96	93	60	76
	37	81	54	89
	58	79	78	88
Fertilizer 2	69	101	56	84
	73	96	61	75
	81	102	69	68

This is a data set on two factors jointly or a two-way table. Sometimes the extra factor is introduced because we are interested in studying its effect also, on its own and in combination with the original factor. At other times, it may be an uncontrollable (like say, the weather) factor which nonetheless has a significant effect on the response, alone or in combination with the original factor.

The following table of means and the chart, called the INTERACTION CHART are quite revealing. The table may be interpreted as follows: As found earlier and as is noted from the last row of this table, Variety 2 gives the most yield. However, under Fertilizer 1, Variety 4 is just as good. Fertilizer 2 on the whole results in higher yield, but only for Varieties 1 and 2, and not for Varieties 3 and 4. So you cannot declare any one variety and any one fertilizer as the best; it depends on which choice of the other factor you use. This situation is called INTERACTION between two factors. The picture does not very clearly show any evidence of interaction.

	Variety				
	1	2	3	4	
Fertilizer 1	63.67	84.33	64	84.33	74.08
Fertilizer 2	74.33	99.67	62	75.67	78.91
	69	92	63	80	

Interaction Chart



How do you analyze data of this kind? You can start by considering the 8 Variety-Fertilizer combinations as one-way classified data and produce an ANOVA of the following kind:

Source	SS	DF	MS	F-ratio	p-value
Var-Fer	3582	7	511.71	3.113	0.0284
Residual	2630	16	164.37		

which shows a certain amount of differences in the eight means. Now just like you decomposed the one-way ANOVA table into three ‘statistically’ independent components, you could decompose the 7 DF SS into 'statistically' independent components as due to Variety (3 DF), Fertilizer (1 DF) and Variety-Fertilizer Interaction (3 DF) as follows:

Source	SS	DF	MS	F-ratio	p-value
Variety	2940	3	980	5.962	0.0063
Fertilizer	88.17	1	88.17	0.536	0.4745
Interaction	553.83	3	184.61	1.123	0.3692
Var x Fer	3582	7	511.71	3.113	0.0284
Residual	2630	16	164.37		

This shows that the interaction (the differences of the difference between fertilizers for the four treatments) is not significant (which was not very clear from the chart) and so it is reasonable to make overall statements about fertilizer comparisons without being specific about the variety, and similarly about variety comparisons without being specific about the fertilizer. You could not do it if the interaction is significant; in that case comparisons can be made of the levels of one factor only for specific levels of the other factor. In this case it turns out that the two fertilizers are not significantly different; the four varieties are significantly different and we have already investigated this from the same data.

Interaction

What exactly is this interaction? It answers the question: Are the varietal differences of the same magnitude and direction irrespective of the fertilizer used? The same question can also be asked as: Are the fertilizer differences of the same magnitude and direction irrespective of the variety used? Both have the same answer. They are both answered by examining the following differences computed from the table of means above:

	Variety			
	1	2	3	4
Fertilizer 1 – Fertilizer 2	-10.66	-15.34	2.00	8.66

Each of these is computed from 6 observations with a variance of $\sigma^2/6$, and a 3 DF SS for Interaction is computed in a manner indicated in the discussions above, to examine if these four quantities are significantly different. If they are, there is interaction; if they are not, interaction is absent. The Interaction Chart is a visual representation of the interaction. It represents the yield values for the four varieties for each fertilizer and joins them by lines, using different colours for different fertilizers. If the lines are parallel, then there is no interaction; if the lines intersect there is interaction. In this graph there is no clear-cut evidence of interaction; there is some, but not significant.

8. Analysis of Covariance

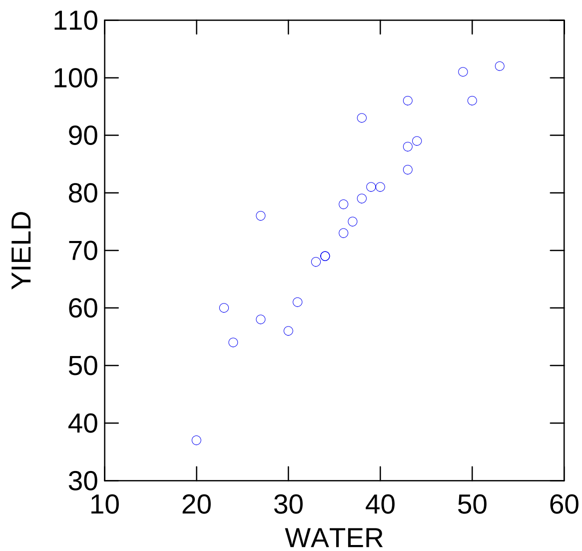
In the foregoing data, we found that the two fertilizers were used equal number of times in each variety and since there was no interaction, the effect of fertilizer was additive and so cancelled out in comparing the varieties. Had there been interaction, then you would have compared the varieties separately for Fertilizer 1 and for Fertilizer 2 and may have come to different conclusions about the comparative average yields of the varieties. If the fertilizers had been used in different numbers you would have adjusted for this differential while comparing treatments.

Suppose the 24 farms involved in this experiment had used different amounts of water indicated by the second number in the data set below.

(Yield, Water)

Variety1	Variety 2	Variety 3	Variety 4
96, 43	93, 38	60, 23	76, 27
37 , 20	81, 39	54, 24	89, 44
58, 27	79, 38	78, 36	88, 43
69, 34	101, 49	56, 30	84, 43
73, 36	96, 50	61, 31	75, 37
81, 40	102, 53	69, 34	68, 33

How would you make the comparison of the varieties? You would then study how the amount of water relates to the yield. To see this, let us make a scatter plot of yield against water:



It appears that in this range the amount of water has an increasing effect on yield and in a linear way. What you would do in this situation is to compute the linear regression of Yield on Water assuming that the regression coefficient is the same in all the varieties, and eliminate the effect of regression. What this means is that you should modify each observation to what it would be if the water used is the same for all observations. If β is the common regression coefficient, then you compute the contribution of water to the variation in yield as an SS with 1 DF because the parameter is the regression coefficient β and the contribution of the varietal differences after eliminating the water effect, is given by an SS with 3 DF. This Analysis of Variance called ANALYSIS OF COVARIANCE is presented here. In this case, most of the observed differences in the varieties are due to the differential water quantities used and so the left-over varietal differences turn out to be insignificant.

The computations can be carried out as follows: You first treat the bivariate (yield, water) data to a procedure similar to ANOVA, where you get SS for yield, SS for water and Sum of Products (SP) for (yield, water) as follows:

	SS Yield	SP Yield x Water	SS Water
Treatment	2940	1466	734
Residual	3272	1544	971
Total	6212	3010	1705

Using this you compute sum of squares due to regression from the Total and Residual rows as:

Regression: $(1544)^2$ /971 = 2455.135 is the Regression SS from the Residual row above, since this eliminates Treatment effect in the regression computation. So the modified residual is: 3272-2455.135 = 816.865

Similarly the modified Total SS = 6212 – $(3010)^2$ /1705 = 899.097, from which the regression-corrected Variety SS to be 899.097 – 816.865 = 82.232. Note that an estimate of the regression coefficient is given by 1544/971 = 1.59 from the Residual row of the above ANOVA. These computations are represented in the ANOVA as follows:

Analysis of Covariance

Source	SS	DF	MS	F-ratio	p-value
VARIETY	82.332	3	27.444	0.638	0.600
WATER	2455.135	1	2455.135	57.106	0.000
Residual	816.865	19	42.993		

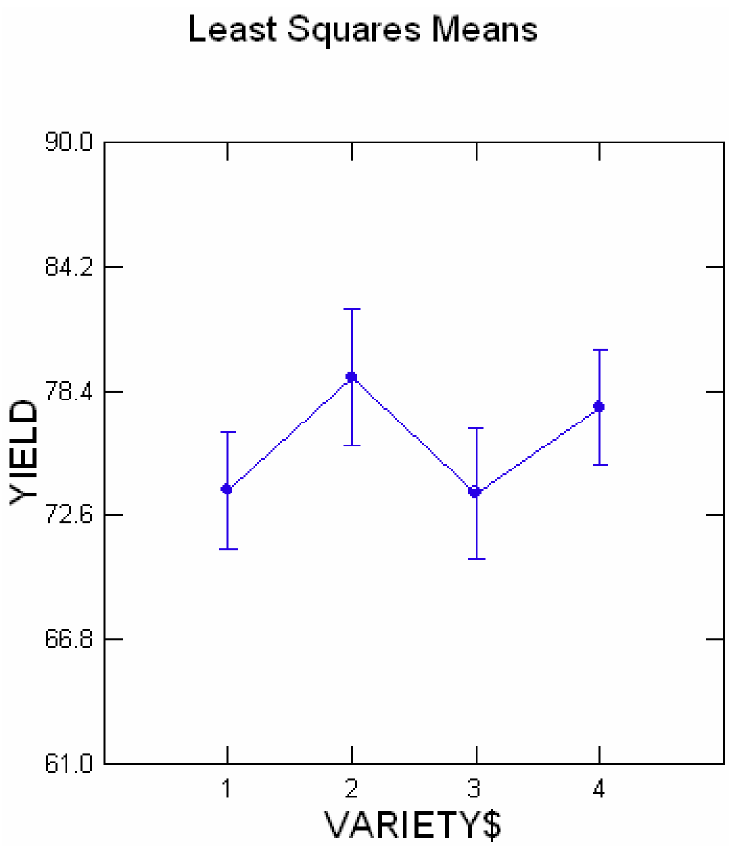
Now, the original means of the yield for the four variety are adjusted for differential water levels by the formula:

Adjusted yield for a variety = original variety mean – regression coefficient * (original water mean for the variety – overall water mean)

Thus for variety 1, it is: 69-1.59*(33.333-36.33) = 73.77 and similarly for other varieties as given in the table below.

Least Squares Means				
Factor	Level	LS Mean	Standard Error	N
VARIETY$	1	73.770	2.750	6.000
VARIETY$	2	79.014	3.181	6.000
VARIETY$	3	73.601	3.022	6.000
VARIETY$	4	77.615	2.695	6.000

* Means are computed after adjusting covariate effect.



The Water test shows that the linear effect of Water on Yield is highly significant (p-value = 0.000) and so it was indeed advisable to have removed its effect in the comparison of Varieties. Once the water effect is removed you notice that it was indeed advisable to have removed its effect in the comparison of Varieties. Once the water effect is removed you notice that the variety effect is not significant (0.600). Compare this to the ANOVA table where you compared Varieties without taking Water into account when the Varietal differences turned out to be quite significant (p-value = 0.004). Thus the differences in varieties happen to be almost completely due to the differential the amount of water used.

9. Multivariate Analysis of Variance (MANOVA)

When you have several response variables, you would like to compare groups based on all the variables. For instance, a famous data set due to Fisher (given in the Appendix) gives four measurements (sepal length, sepal width, petal length, petal width) on 50 specimens from each of three species of the Iris plant---iris setosa, iris versicolor, iris virginica.

Suppose you are interested in investigating the differences between the three species and in particular, to see if the species are distinguishable by using all the four measurements *together*. A joint analysis of the four measurements gives much more information than the individual measurements. The relationships among the measurements is also an important item of information. Hence it is advisable to carry out a *multivariate* analysis of the data.

In terms of these multiple variables, the first question we would like to be answered is whether the three species differ in terms of the average of the multiple measurements and if so by how much. Let us first compute these average (arithmetic mean) measurements for the three species:

Results for SPECIES = Setosa
Number of Observations: 50

Means			
SEPALLEN	SEPALWID	PETALLEN	PETALWID
5.006	3.428	1.462	0.246

Results for SPECIES = Versicolor
Number of Observations: 50

Means			
SEPALLEN	SEPALWID	PETALLEN	PETALWID
5.936	2.770	4.260	1.326

Results for SPECIES = Virginica
Number of Observations: 50

Means			
SEPALLEN	SEPALWID	PETALLEN	PETALWID
6.588	2.974	5.552	2.026

Results for all species:
Number of Observations: 150

Means			
SEPALLEN	SEPALWID	PETALLEN	PETALWID
5.843	3.057	3.758	1.199

To answer questions on the vector of means, as in univariate analysis, we make some assumptions (which should be tested). Now besides the variances of the four variables, we should also consider the covariances (or correlations) between pairs of variables. The variances and covariances together form the Covariance Matrix (also known as Variance-Covariance Matrix or Dispersion Matrix). This is a symmetric matrix, the diagonal elements of which are the variances and the off-diagonal elements the covariances. We make assumptions analogous to univariate analysis: independence of the observations; normality, which in this case is multivariate normality (a generalization of the normal distribution to the multivariate case), and homoscdasticity (which in this case is the equality of the covariance matrices across the species).

Just as we began with the total sum of squares in the univariate case, we begin with the TOTAL SUM OF SQUARES AND CROSS PRODUCTS MATRIX in the multivariate case. Just as we decomposed the total sum of squares into suitable components in the Analysis of Variance, we decompose the Total Sum of Squares and Cross Products Matrix (SSCP) into suitable components in the multivariate case.

This Total SSCP matrix is:

Total Sum of Squares and Cross Products				
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	102.2			
SEPALWID	-6.323	28.307		
PETALLEN	189.9	-49.119	464.3	
PETALWID	76.924	-18.124	193.0	86.570

This has two components: SSCP due to Species; SSCP due to Residual. The purpose of the Multivariate Analysis of Variance (MANOVA) is to effect a decomposition of this matrix due to the two sources.

If you want to compute the covariance matrix from the above, divide it by 149 (n-1), just as you did in the univariate case. But that is not a correct estimate of the common covariance matrix, since the total SSCP arises from a combination of data from three species with three possibly different mean structures. For what it is worth, here are the covariance and correlation matrices from the total SSCP matrix.

Covariance Matrix				
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	0.686			
SEPALWID	-0.042	0.190		
PETALLEN	1.274	-0.330	3.116	
PETALWID	0.516	-0.122	1.296	0.581

Pearson Correlation Matrix				
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	1.000			
SEPALWID	-0.118	1.000		
PETALLEN	0.872	-0.428	1.000	
PETALWID	0.818	-0.366	0.963	1.000

The decomposition is similar to the ANOVA decomposition, the difference being that the sums of squares are now replaced by SSCP matrices. In these matrices the diagonals are sums of squares of individual variables, and they are to be computed in the same manner as for ANOVA in respect of that variable. An off diagonal element of these matrices correspond to a pair of variables, say, sepallen and sepalwid, and it is a cross-product of two terms one from sepallen (the same term which was squared for a sepallen sum of squares) and another from sepalwid (the same term which was squared for a sepalwid sum of squares). So if you know how to compute a sum of squares for ANOVA, you can imitate that step for each element of the corresponding SSCP matrix. Thus you can compute the analogue of ANOVA, where the SS are now SSCP. This forms a part of the MANOVA table.

Corresponding to one-way ANOVA, here for the Iris data, we have a one-way MANOVA. The degrees of freedom is the same as that for the corresponding SS in ANOVA.

Species Sum of Product Matrix $H = B'A'(A(X'X)^{-1}A')^{-1}AB$				
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	62.568			
SEPALWID	-17.956	5.153		
PETALLEN	1.618E+002	-46.421	4.182E+002	
PETALWID	70.399	-20.203	1.820E+002	79.210

Residual Sum of Product Matrix G = E'E				
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	39.600			
SEPALWID	11.633	23.154		
PETALLEN	28.113	-2.697	46.123	
PETALWID	6.525	2.079	11.041	7.360

Total Sum of Squares and Cross Products				
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	102.2			
SEPALWID	-6.323	28.307		
PETALLEN	189.9	-49.119	464.3	
PETALWID	76.924	-18.124	193.0	86.570

Notice that the first two matrices add up to the third just like in ANOVA. The Residual SSCP matrix divided by the Residual degrees of freedom is an estimate of the common covariance matrix Σ just like the Residual Mean Square gave us an estimate of the common variance in ANOVA.

In ANOVA, you computed the ratio of Hypothesis mean square to Residual mean square as an F-statistic. What is the multivariate analogue? What is the ratio of two matrices? There are different answers to these questions. Based on the particular answer, the analogue of the F-ratio varies. One could take the ratio of the determinants of the two matrices (essentially what is called the Wilks's lambda), or the largest eigenvalue of the product of the numerator and the inverse of the denominator, or the trace (sum of eigenvalues or equivalently the sum of the diagonal values) of this matrix, etc. A few standard statistics like these have been suggested and in simple and balanced cases they all turn out to the same as shown below for the iris data. The F-distribution here is an approximation in general and its degrees of freedom in this case is (2 times (the number od species-1), no of observations-number of variables-1):

Multivariate Test Statistics				
Statistic	Value	F-ratio	df	p-value
Wilks's Lambda	0.043	7.998E+002	4, 145	0.000
Pillai Trace	0.957	7.998E+002	4, 145	0.000
Hotelling-Lawley Trace	22.064	7.998E+002	4, 145	0.000

All these tests show that the three species mean vectors are very significantly different. This is MANOVA.

For most MANOVA problems, you could extend your ANOVA ideas in this manner by substituting the SS by SSCP in the manner explained. For complex and/or unbalanced models, the three statistics and their distributions could be different.

Reference:

O.J.Dunn and V.A.Clark (1987). Applied Statistics: Analysis of Variance and Regression, 2nd Edition. New York: John Wiley & Sons.

R.Miller (1985). Multiple comparisons. Kotz, S. and Johnson, N. L.,eds.

Encyclopedia of Statistical Sciences, vol. 5. New York: John Wiley & Sons, 679–689.

H.Sahai and M.I.Ageel (2000): The Analysis of Variance: Fixed, Random and Mixed Models. Boston: Birkhauser.

Iris Data:

The first four columns are sepal length, sepal width, petal length, petal width of setosa, the next four of versicolor and the last four virginica:

5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
4.3	3.0	1.1	0.1	6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2

5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
5.1	3.7	1.5	0.4	6.1	2.8	4.0	1.3	5.6	2.8	4.9	2.0
4.6	3.6	1.0	0.2	6.3	2.5	4.9	1.5	7.7	2.8	6.7	2.0
5.1	3.3	1.7	0.5	6.1	2.8	4.7	1.2	6.3	2.7	4.9	1.8
4.8	3.4	1.9	0.2	6.4	2.9	4.3	1.3	6.7	3.3	5.7	2.1
5.0	3.0	1.6	0.2	6.6	3.0	4.4	1.4	7.2	3.2	6.0	1.8
5.0	3.4	1.6	0.4	6.8	2.8	4.8	1.4	6.2	2.8	4.8	1.8
5.2	3.5	1.5	0.2	6.7	3.0	5.0	1.7	6.1	3.0	4.9	1.8
5.2	3.4	1.4	0.2	6.0	2.9	4.5	1.5	6.4	2.8	5.6	2.1
4.7	3.2	1.6	0.2	5.7	2.6	3.5	1.0	7.2	3.0	5.8	1.6
4.8	3.1	1.6	0.2	5.5	2.4	3.8	1.1	7.4	2.8	6.1	1.9
5.4	3.4	1.5	0.4	5.5	2.4	3.7	1.0	7.9	3.8	6.4	2.0
5.2	4.1	1.5	0.1	5.8	2.7	3.9	1.2	6.4	2.8	5.6	2.2
5.5	4.2	1.4	0.2	6.0	2.7	5.1	1.6	6.3	2.8	5.1	1.5
4.9	3.1	1.5	0.2	5.4	3.0	4.5	1.5	6.1	2.6	5.6	1.4
5.0	3.2	1.2	0.2	6.0	3.4	4.5	1.6	7.7	3.0	6.1	2.3
5.5	3.5	1.3	0.2	6.7	3.1	4.7	1.5	6.3	3.4	5.6	2.4
4.9	3.6	1.4	0.1	6.3	2.3	4.4	1.3	6.4	3.1	5.5	1.8
4.4	3.0	1.3	0.2	5.6	3.0	4.1	1.3	6.0	3.0	4.8	1.8
5.1	3.4	1.5	0.2	5.5	2.5	4.0	1.3	6.9	3.1	5.4	2.1
5.0	3.5	1.3	0.3	5.5	2.6	4.4	1.2	6.7	3.1	5.6	2.4
4.5	2.3	1.3	0.3	6.1	3.0	4.6	1.4	6.9	3.1	5.1	2.3
4.4	3.2	1.3	0.2	5.8	2.6	4.0	1.2	5.8	2.7	5.1	1.9
5.0	3.5	1.6	0.6	5.0	2.3	3.3	1.0	6.8	3.2	5.9	2.3
5.1	3.8	1.9	0.4	5.6	2.7	4.2	1.3	6.7	3.3	5.7	2.5
4.8	3.0	1.4	0.3	5.7	3.0	4.2	1.2	6.7	3.0	5.2	2.3
5.1	3.8	1.6	0.2	5.7	2.9	4.2	1.3	6.3	2.5	5.0	1.9
4.6	3.2	1.4	0.2	6.2	2.9	4.3	1.3	6.5	3.0	5.2	2.0
5.3	3.7	1.5	0.2	5.1	2.5	3.0	1.1	6.2	3.4	5.4	2.3
5.0	3.3	1.4	0.2	5.7	2.8	4.1	1.3	5.9	3.0	5.1	1.8