Follow-up Test to a One Factor Analysis of Variance

The material above showed techniques on how to test the overall null hypothesis (in a one factor design) of equal population means in several populations. The basic steps involved finding the F ratio and then comparing it to the appropriate critcal F value. If the F ratio was less than the critical value, you retain the null hypothesis. In that case, your analysis is complete. However, if the F ratio was equal to or larger than the critical value, you would reject the null hypothesis. But, if you **REJECT THE NULL HYPOTHESIS**, then the further question remains: exactly where do the differences in means occur? Are all of the population means different? Or, if it was a 3 group study like the first example above, perhaps the means are different in populations 1 and either of 2 and 3, but 2 and 3 are equal to one another. Thus, if one rejects the null hypothesis, there is more work to do to better pinpoint where the differences are. This is the subject of **FOLLOW-UP** tests and we now turn our attention to that in the context of a one factor analysis of variance problem. There are extensions of these tests to multi-factor designs, which are discussed next, but these will not be discussed here.

What if we examine the data from the first experiment discussed above where there were 3 different methods for assembling electronic components. These data were presented on page 280. Recall the descriptive statistics.

	N	MEAN	STDEV
Dia	6	23.33	3.14
Help	6	20.833	2.317
CAD	6	30.00	3.52

Since we wanted to see if there were differences in the population means, we completed a one factor analysis of variance and the summary table is below.

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F	p
METHOD	2	269.44	134.72	14.63	0.000
ERROR	15	138.17	9.21		
TOTAL	17	407.61			

From the analysis of variance output, we rejected the null hypothesis and now must try to pinpoint exactly where the population mean differences lie. To better isolate where the differences are, I will outline what is called the **TUKEY FOLLOWUP TEST**. This is merely one of many procedures that you could use to do this "follow-up" analysis but, it is the most straight forward and will satisfactorily meet our needs. Here are the steps you need to follow with this procedure.

1. First, you **order the means** from left to right, from the largest to the smallest. For

our data, it looks as follows.

CAD	Dia	Help
30	23.333	20.833

2. Second, you will need to **calculate a measure of "error"** that will be used (like in the t test) as a denominator for our test statistic. The formula is as follows.

Standard Error = Sqrt [MS WG / n (for a group)]

The MS WG term comes directly from the analysis of variance summary table and for our table above is 9.21. The n represents the number of people in each group and is 6 in this case. Note: Tukey tests work best when there are equal ns in the groups. For our data therefore, the standard error would be as follows.

Standard Error = Sqrt [9.21 / 6] = Sqrt [1.535] = **1.238**

3. The next step is to form Q values which, in the literature, are referred to as "Studentized Range Statistics". The Q formula looks as follows and is <u>similar</u> to the t test for differences in means.

$$Q = (Mean_L - Mean_S) / \text{Standard Error}$$

With the ordered list of means from 1 above, you first select the <u>pair of means</u> that produces the <u>largest</u> difference since we <u>must</u> find a significant difference there. In the Q formula, we put on the left side in the numerator the larger of the two means (subscript L means larger) and the smaller of the two means on the right side. This means that the numerator of the Q test will always be positive (although it could be 0 if the last two means happen to be equal). Because of this formula set-up, the Q values cannot be negative and therefore are not exactly like t values that can take on both positive and negative values. Since there are 3 possible paired comparisons in this case, we will need 3 Q values: one for each comparison.

First Q = (30 - 20.833) / 1.238 = **7.399**
Second Q = (30 - 23.333) / 1.238 = **5.381**
Third Q = (23.333 - 20.833) / 1.238 = **2.017**

5. As with any statistical test, we need to compare our calculated values (Q's in this case) with some critical value (critical Q value). If the **calculated Q is equal to or larger than the critical value, we reject the null hypothesis**. But, if the calculated Q value is less than the critical value, we retain. The question is, where do we get the critical value of Q? The tables on the next 2 pages show the .05 critical values of Q. To enter the table, start with the .05 part at the top. The #G at the top means the number of groups in the study; it is 3 for our first example. The df on

the side represents the number of degrees of freedom for the **WITHIN OR ERROR** term of the analysis of variance table; 15 for our example. For this example, you need to find the intersection of the 3rd column and the 15 row and luckily there happens to be one: 3.674. **THIS IS OUR CRITICAL Q VALUE** in this study.

6. If you compare the calculated Q values from 4 to the critical value based on using the table as described in 5, you will see that the calculated Q values between 3 and 2 (7.399) and between 3 and 1 (5.38116) are **LARGER** than the critical value. Therefore, we reject these 2 separate null hypotheses. However, the comparison between 2 and 1 did not turn out to be significant. The bottom line in this case is that we think that population mean 3 is different than population mean 2 and population mean 1 but that we do not think that the means from populations 1 and 2 are different. In this study, it appears that the computer-aided method is better than either of the other two.

STUDENTIZED RANGE (Q) VALUES

.05 alpha level: df = df for MS(WG); #G = number of groups

df\#G	2	3	4	5
1	17.97	26.98	32.82	37.08
2	6.085	8.331	9.798	10.88
3	4.501	5.910	6.825	7.502
4	3.927	5.040	5.757	6.287
5	3.635	4.602	5.218	5.673
6	3.461	4.339	4.896	5.305
7	3.344	4.165	4.681	5.060
8	3.261	4.041	4.529	4.886
9	3.199	3.949	4.415	4.756
10	3.151	3.877	4.327	4.654
11	3.113	3.820	4.256	4.574
12	3.082	3.773	4.199	4.508
13	3.055	3.735	4.151	4.453
14	3.033	3.702	4.111	4.407

15	3.014	3.674	4.076	4.367
16	2.998	3.469	4.046	4.333
17	2.984	3.628	4.020	4.303
18	2.971	3.609	3.997	4.277
19	2.960	3.593	3.977	4.253
20	2.950	3.578	3.958	4.232
24	2.919	3.532	3.901	4.166
30	2.888	3.486	3.845	4.102
40	2.858	3.442	3.791	4.039
60	2.829	3.399	3.737	3.977
120	2.800	3.356	3.685	3.917

Note 1: Table abridged from Tables of range and studentized range, Annals of Mathematical Statistics, 31(1960), 1122-1147 by Harter. Used by permission of Institute of Mathematical Statistics.

Consider another experiment comparing 4 different methods of reading text material with the following results. This example is different than the 4 group study in the one factor analysis of variance section above since that one did not lead to a rejection of the null hypothesis. The data are as follows.

DATA FROM A 4 GROUP READING STUDY

Meth1	Meth2	Meth3	Meth4
26	24	32	23
25	17	34	29
29	16	29	26
21	13	19	20
20	21	27	24
18	19	28	25

If you complete the one factor analysis of variance, you would obtain the following analysis of variance summary table results.

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F	p
METHODS	3	297.5	99.2	5.79	0.005
ERROR	20	342.5	17.1		
TOTAL	23	640.0			

LEVEL	n	MEAN	STDEV
Meth1	6	23.167	4.167
Meth2	6	18.333	3.882
Meth3	6	28.167	5.193
Meth4	6	24.500	3.017

From the analysis of variance summary table, the F ratio of 5.79 and the p value (less than .05) allow us to reject the overall null hypothesis. Thus, we conclude that there are some differences in the population means. But, again, where exactly are these differences? The Tukey test will help. Again, we need to follow the steps from the first example. See below.

1. Order the means. The "method group" is below the mean.

28.167	**24.5**	**23.167**	**18.333**
Meth3	Meth4	Meth1	Meth2

2. The standard error is found as follows.

   Standard Error = Sqrt [17.1 / 6] = Sqrt [2.85] = **1.688**

3. We need to calculate all the Q values starting with the largest difference between 2 means and working our way down to the smallest difference in 2 means.

   Q1 = (28.167 - 18.333) / 1.688 = **5.82516** <--- Compares 3 and 2
   Q2 = (24.5 - 18.333) / 1.688 = **3.65302** <--- Compares 4 and 2
   Q3 = (28.167 - 23.167) / 1.688 = **2.96174** <--- Compares 3 and 1
   Q4 = (23.167 - 18.333) / 1.688 = **2.86341** <--- Compares 1 and 2
   Q5 = (28.167 - 24.5) / 1.688 = **2.17214** <--- Compares 3 and 4
   Q6 = (24.5 - 23.167) / 1.688 = **0.789601** <--- Compares 4 and 1

4. The Q values range from 5.8 to .8. We now need to find the critical Q value from the table and the #G value will be 4 and the degrees of freedom value will be 20. From the table, the .05 value is 3.958.

5. Comparing the calculated values in 3 to the critical value in 4, only the first Q value of 5.82516 is larger than the critical value. Therefore, we conclude that there is only

a difference between population means 2 and 3 (methods 2 and 3). This also implies that method 4 is equal to method 2, method 1 is equal to method 3, method 1 is equal to method 4, method 2 is equal to method 4, and method 3 is equal to method 4. Of the 6 possible comparisons, only one appears to be significant.

Notes on Follow-up Tests

1. The need to conduct follow-up tests only exists if and when you reject the null hypothesis with your F ratio from the analysis of variance. In the literature, this is sometimes called the "omnibus" F test since you are, with one whack, testing to see if there are any differences in population means. If you retain, the analysis is over. Only in the case that you reject will you need to do follow-up tests.

2. The test illustrated above is called the Tukey test. In the literature, the Tukey test is called a "conservative" test. Recall from above that I said that if you reject the null hypothesis with the F ratio, that the largest difference in means will necessarily result in a significant Q value in your follow-up test. Well, I fudged the truth a little! Under certain circumstances, an F ratio that is **JUST BARELY SIGNIFICANT** (that is, just big enough to reject) will not always show a significant Q value in the Tukey follow-up test. This would be confusing to say the least since the ominbus F test says there are differences and the Tukey follow-up tests says there are not. There are other follow-up tests where this problem will not be encountered. While other follow-up tests such as Scheffe and Newman-Keuls will not be discussed here, the interested reader should consult a more advanced book to obtain information on these alternative tests (see Glass and Hopkins, 1984 for example).

3. As was mentioned at the beginning of this section, follow-up tests apply not only to one factor analysis designs but also to multi-factor designs. What has been discussed above only presents the simplest case scenario as an introduction to follow-up techniques. Again, consult more advanced books for additional information.

Two Factor Analysis of Variance

In the one factor section above, I outlined the steps involved in using ANOVA to test a null hypothesis of the equality of several population means in a one independent variable experiment or study. Suppose that you wanted to do experiments to see the impacts of type of instructional method on performance, and type of feedback on performance. One way to do this is to first conduct one experiment manipulating the first independent variable and see what happens. Then, you could conduct the second experiment and see what happens. By examining the results from both studies, you will get some idea about the impacts of these two independent variables.

However, there are some disadvantages with the above approach. First of all, you will need to do two separate experiments and doing two is obviously more time consuming and resource depleting than doing only one investigation. Since it is becoming increasingly difficult to do experiments at all, the more you have to do the worse off you probably will be. Secondly, one of the problems of doing two separate studies is that it is possible that the two independent variables work together in some special way such that using one particular method of instruction coupled with one particular feedback strategy might produce an overall benefit in performance that would not be evidenced by another combination of teaching method and feedback method. However, if you do two separate studies, there is no way to discover such a possibility since you are only manipulating one independent variable at a time. Thus, there is one type of information that is not available when you manipulate two independent variables, each one in its own separate experiment. We call this special type of information an **INTERACTION**; to be discussed shortly.

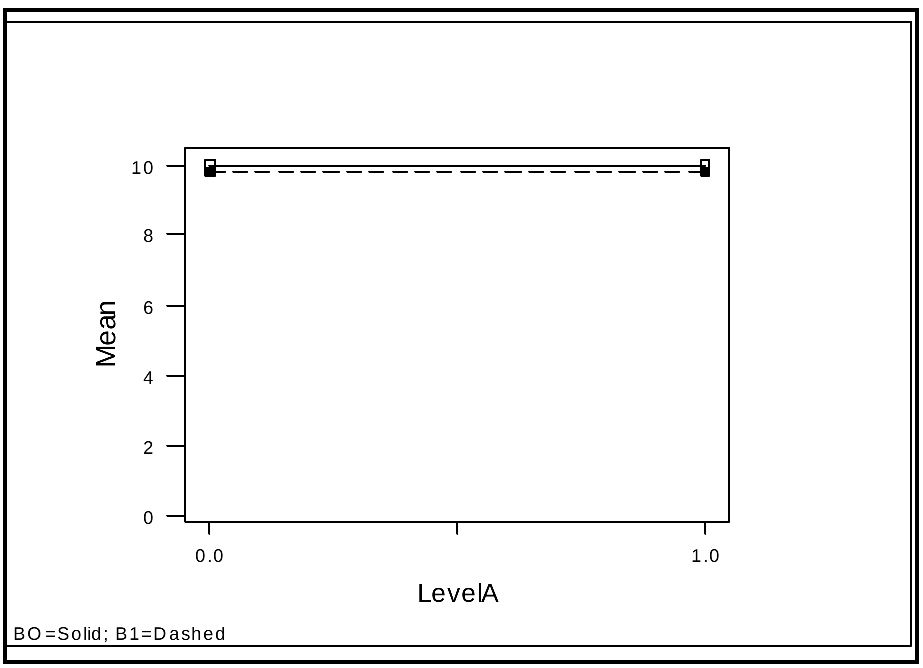
What we need as a solution to both of the above problems is a way to conduct one study and manipulate both independent variables at the same time. Many studies are like that and the name for such an investigation is called a **TWO FACTOR DESIGN**. In analysis of variance, a two factor ANOVA is simply where two independent variables have been manipulated in one investigation and the results have been analyzed by analysis of variance techniques. I will present a fairly brief overview to the two factor ANOVA and also, since the analysis becomes more complicated, show how Minitab can complete this analysis for us. I also want to introduce the concept of an interaction. Before moving on to a concrete data analysis example, I want to illustrate how the data can be summarized over the two factors in table form and in graph form.

What if you had a two factor design with Factor A and Factor B and the table of cell means looks like the following. The cell means have been slightly highlighted.

		Level of Factor B		**Means**
		0	1	
Level of Factor A	0	10	10	10
	1	10	10	10
Means		10	10	10

Factor A has two levels: 0 and 1, and Factor B has two levels: 0 and 1. Assume that some people have been randomly assigned to each of the cells and the **mean performance** for each cell is indicated (all 10's in this case). The row means of 10 for both levels 0 and 1 on Factor A indicate that performance is the same for level 0 and level 1. The independent variable A seems not to make any difference. The same thing applies to Factor B in that the mean down the first column (level 0 on Factor B) is the same as the mean down the second column (level 1 on Factor B). Again, independent variable B seems to have no impact. What would a graph of these data look like? Well, look at the top of the next page for that illustration. The way this graph is drawn is as follows. First, I let the baseline be one of the two independent variables or factors; in this case, I let the baseline be Factor A. Since there are 2 levels of A, I have put tick marks on the left side for level 0 and on the right side for level 1. The vertical axis will be the dependent variable; mean on the outcome variable (test scores, heights, speed, or whatever). Finally, to indicate Factor B, I have let lines in the graph do that, one for level 0 on Factor B and another line for level 1 on Factor B. The solid line in the graph is for the Factor B level 0 data; at A0 a point is raised to a value of 10 (from the data table) and at A1, another point is raised to a height of 10. Thus, the B0 line is flat. The B1 line (dashed) is the same in this case; at A0 = 10 and at A1 = 10. Although these two lines are in the same exact location, I have separated them slightly just for visual effects.

To interpret the data using the graph, we would look at the **average heights** of the A0 and A1 dots. In both cases on this graph, the two dots at A0 average to 10 and the two dots at A1 also average to 10; thus, we conclude that Factor A has no impact. To examine the impact of B, we average the two dots on each line; for B0 line and for B1 line. Since both dots on both lines average to 10, we conclude that there is no dependent variable impact difference due to B. What we have here is a classic case of "nothing working"!

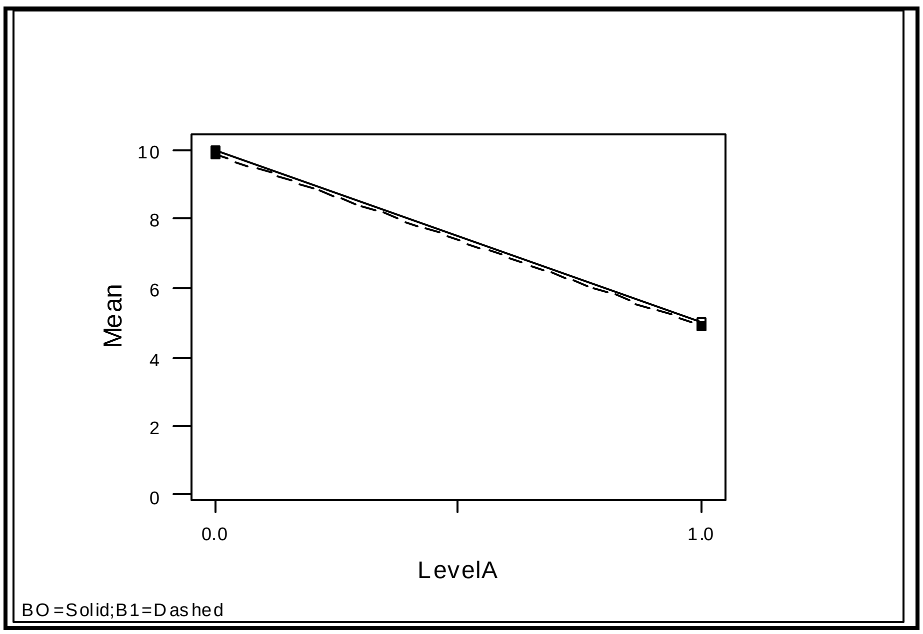


Consider this second example.

		Level of Factor B		Means
		0	1	

Level of Factor A	0	10	10	10
	1	5	5	5
Means		7.5	7.5	7.5

In this case, performance on Factor A level 0 is 10 (row mean) compared to a row mean of 5 on level A1. Thus, since the row means are different, it appears that Factor A has some impact. However, for Factor B, since the column means are the same, it appears that there is no impact based on manipulating independent variable B. As a graph, it would look like the following.



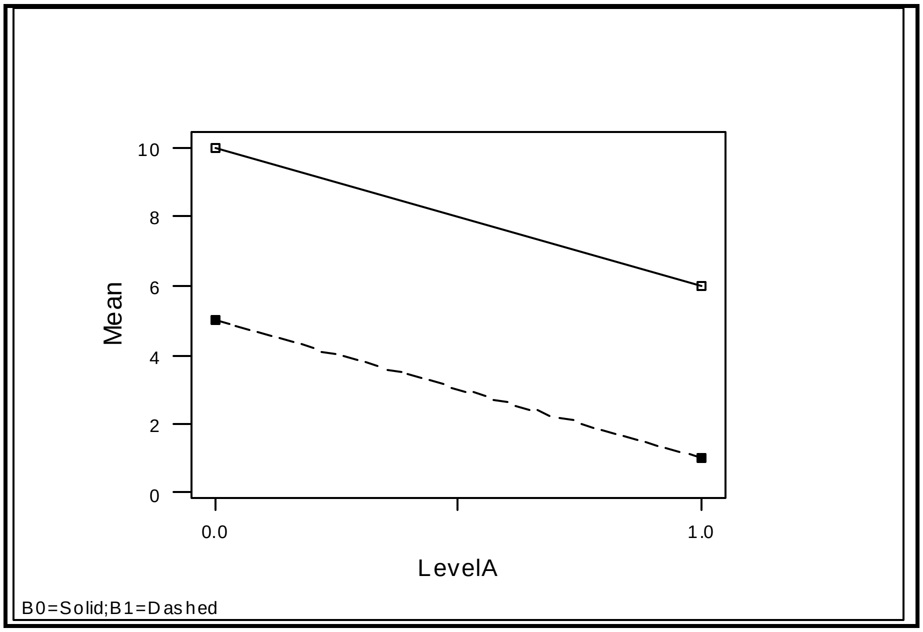
The different levels of the dots above A0 and A1 indicate that Factor A has an impact with A0 being better, but since the two B lines (B0 and B1) are the same, it appears that Factor B has no effect. Thus, one of the two independent variables seems to have an impact.

Look at a third example.

		Level on Factor B		**Means**
		0	1	
Level on Factor A	0	10	5	7.5

	1	6	1	3.5
Means		8	3	3.5

Here, both the row means (A0 and A1) are different and the column means are different (B0 and B1) so it appears that <u>both</u> Factors A and B have an effect. Look at the graph at the top of the next page.



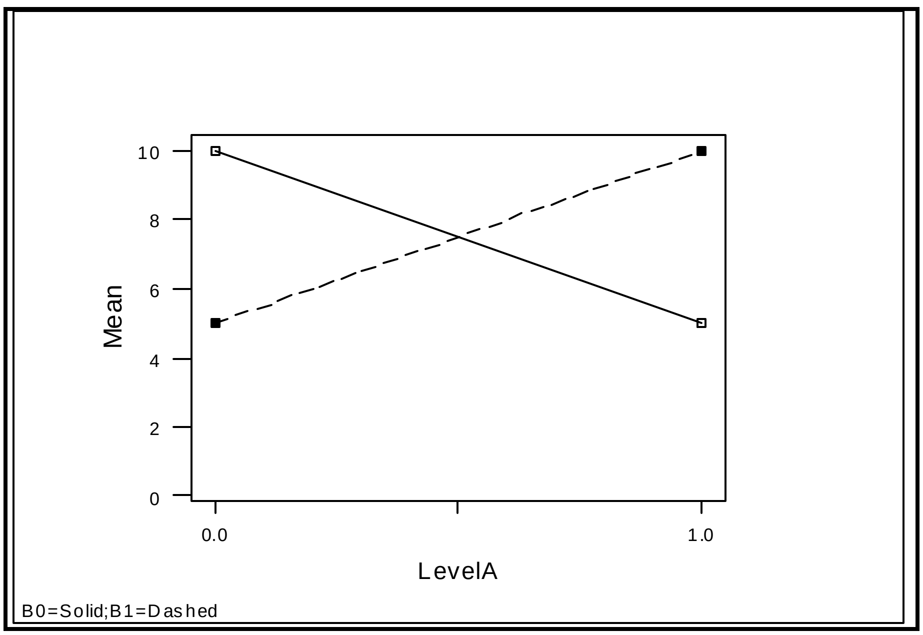
Here, the general heights of the dots above A0 are higher than for A1 (averages of 7.5 and 3.5 respectively) and therefore there seems to be some impact due to Factor A. Also, the lines for B0 and B1 (averages of 8 and 3 respectively) seem to be at different levels (with the B0 line being higher) so there appears to be some impact due to Factor B.

Consider this final example.

		Level of Factor B		**Means**
		0	1	
Level on Factor A	0	10	5	7.5
	1	5	10	7.5

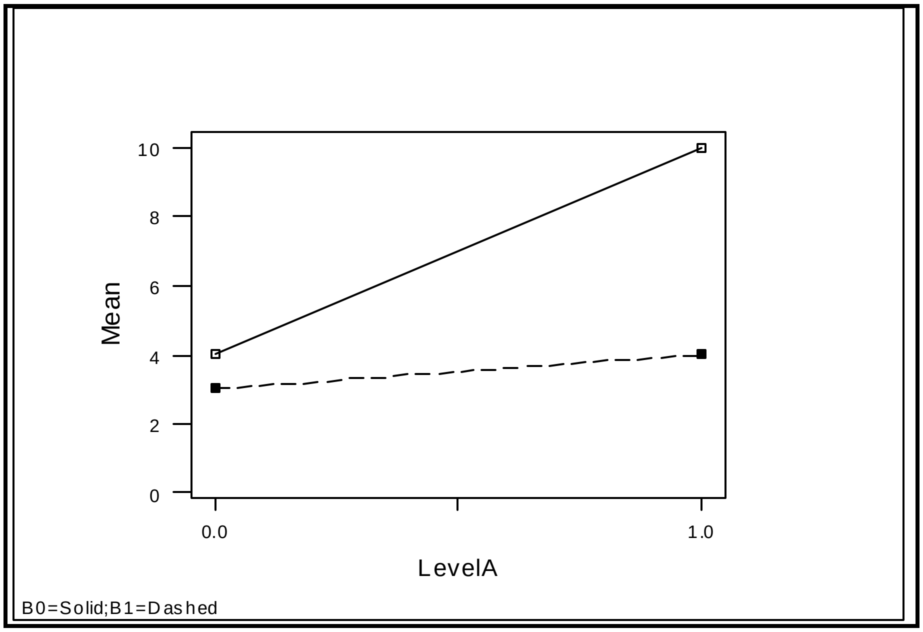
Means	7.5	7.5	7.5

Note from the table that there are no differences between the row means (both 7.5) indicating that there is no impact due to Factor A, and there are no differences between the column means (both 7.5) indicating there is no impact due to Factor B. But, look at the graph. Something is going on! Here we see the classic case of what is called an **interaction**. An interaction is something that appears to be going on between the two factors that does not show up on either factor separately. For example, if you are talking about maximizing performance, then the best performance occurs either when method A1 is combined with B1, or when method A0 is combined with B0. If you happen to be in the A1 B0 or A0 B1 combination, performance is much lower. So, which level on Factor A you should be assigned to or which level of Factor B should you be assigned to? It all depends! To answer which level on A you should be assigned to maximize your performance depends on which level of B you will have. Therefore, stating what happens in terms of producing better performance cannot be done unless you take both independent variables into account. This is what an interaction means: which level of A is best depends upon which level of B you are at. Here is the graph of the data.



This is the famous X type interaction pattern. Clearly, the lines are not parallel. In the literature, this type of a pattern is also called **"disordinal"**. Disordinal simply means that the lines cross over at some point. Interpretatively, this means (in this case) that at A0, B0 is best but at A1, B1 is best. So, from A0 to A1, there is a switch in which level of B is better. The graph at the top of the next page shows a different type of an interaction, one where the lines are not parallel but, do not cross over each other.

Notice for this graph, that the lines are not parallel. That is the key to having an interaction present in the data. But, this graph as mentioned above is different in that the lines do not cross one another. In fact, in this case, if you asked which level of B you should be assigned to, it does not matter in one sense in that B0 is always better (the B0 line is always higher than the B1 line), no matter which level of A you happen to be in. However, it should be clear that the difference between B0 and B1 is much smaller at A0 than it is at A1. Thus, we see a differential difference in performance between A0 and A1 clearly dependent on which level of B you are in. While B0 is always best, the best **combination** is being in A1 and B0 together.



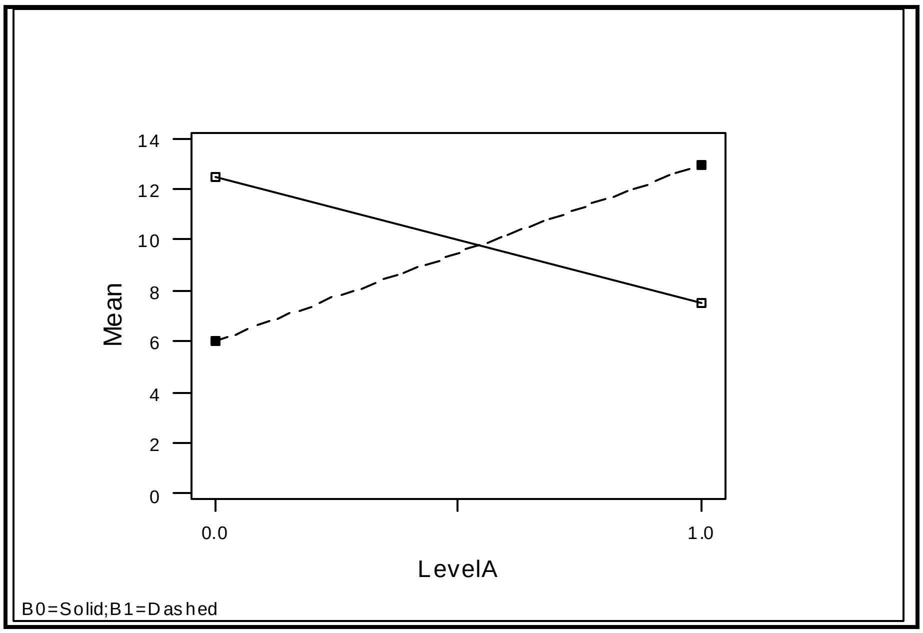
Now let's work through a simple two way or two factor ANOVA problem. Recall in the one factor model, we evaluated one final F ratio to determine whether to retain or reject the null hypothesis. In that type of design, there was only <u>one</u> F ratio. But, in the present case, not only have we added a second independent variable and need to decide whether to retain or reject a null hypothesis related to it, there also is the problem of testing to see if there is an interaction or not. Thus, in the two factor ANOVA, we will go from finding one F ratio and testing one null hypothesis, to finding <u>three F ratios and three null hypotheses</u>.

What if you had what we call a two by two design where there are two independent variables and two levels on each. After the study is completed, you have data on some criterion measure and also you know which particular combination of level on Factor A and level on Factor B each person was in. The data are presented below.

		Level on Factor B		**Means**
		0	1	

Level on Factor A	0	15,16,13,10,10,11: **M=12.5**	7,8,5,8,5,3: **M=6**	**9.25**
	1	9,10,8,5,8,5: **M=7.5**	13,17,14,12,12,10: **M=13**	**10.25**
Means		**10**	**9.5**	**9.75**

Again, the slightly shaded section of the table represents the 4 cell means. I have put the data in each cell plus I have boldfaced the means of the cells (12.5,6,7.5, and 13). Also, note that the row (9.25,10.25) and column (10,9.5) means are bolded at the side and bottom. The value at the lower right corner (9.75) is called the grand mean; ie, the mean of all the data. Here is a graph of the cell means.



From the looks of the graph, it appears that we have one of these interaction situations; in fact, it looks like this is a disordinal pattern. Could this be a case where there are no A and B effects but only the interaction is present? We will see.

The computational procedures for the two factor ANOVA are essentially the same as for the one factor ANOVA. Let's consider the overall plan of attack for our analysis. First, we need to combine the data from all the cells into a stacked (or total) set and then obtain the sum of the squared deviations around the grand mean of 9.75. This will be SS TOT. Then, since the rows represent the "Factor A" effect, we obtain the sum of the squared deviations of <u>each row mean</u> from the overall grand mean (9.25 - 9.75,10.25 - 9.75). This will be our SS ROWS. Then, since the columns represent the "Factor B" effect, we obtain the sum of the squared deviations of <u>each column mean</u> from the overall grand mean (10 - 9.75,9.5 - 9.75). This will be our SS COLS. The next step is to obtain the sum of the squared deviations around the mean of each separate cell and this will be our SS WG. Finally, we can add together the SS ROWS + SS COLS + SS WG terms. The difference between this sum and the SS TOT will be the part attributable to the interaction, SS INTERACTION. The model in a two factor ANOVA is as follows.

SS TOT = SS ROWS + SS COLS + SS INTERACTION + SS WG

The rows will be our A factor, the columns will be our B factor and the interaction will be our test of whether the lines are parallel or not. As usual, the SS WG is our error term. Look at the analysis of variance output from Minitab for this problem.

ANALYSIS OF VARIANCE FOR THE TWO FACTOR DESIGN

Source	DF	SS	MS	F	p
FactA	1	6.000	6.000	1.17	0.293
FactB	1	1.500	1.500	0.29	0.595
A*B	1	216.000	216.000	41.94	0.000
Error	20	103.000	5.150		
Total	23	326.500			

In this type of two by two design, the F ratio for FactA is obtained by dividing the MS for FactA by the ERROR MS (6 / 5.15 = 1.17). The F ratio for FactB is obtained by dividing the MS for FactB by the ERROR MS (1.5 / 5.15 = .29). Finally, for the interaction F ratio, we divide the MS for Interaction by the ERROR MS (216 / 5.15 = 41.94). Note in this case that the MS terms are all divided by the same MS error term. This is not true in all ANOVA designs but it is true in our completely randomized two factor design. To make the decisions about whether to retain or reject each null hypothesis, we need to have a critical 95th percentile rank value from the appropriate F distribution with the appropriate degrees of freedom. In this case, the same F distribution with 1 and 20 degrees of freedom is used for each test. Again, we can let Minitab read off this tabled value for us.

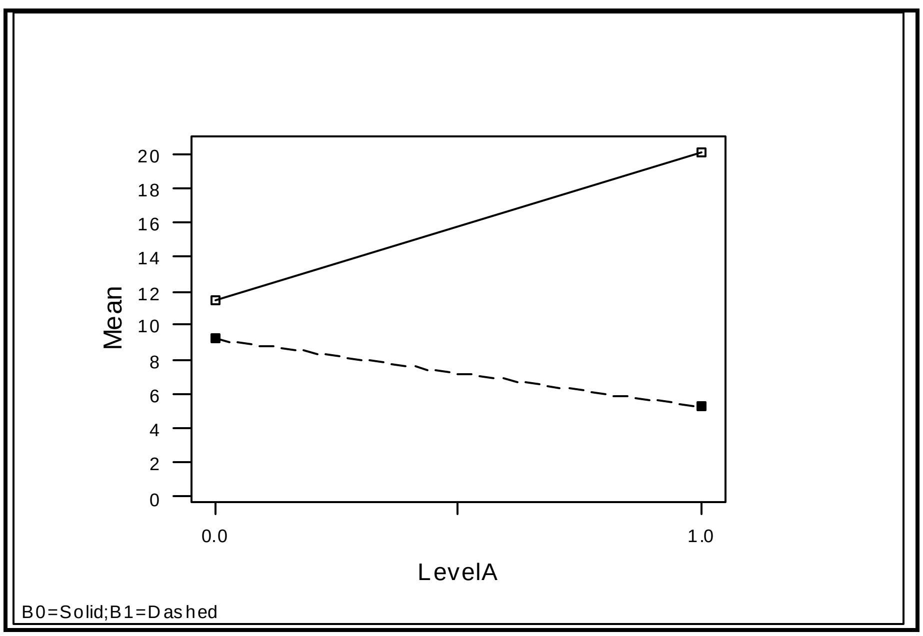
MTB > **invcdf .95;**
SUBC> **f 1 20.**
  .9500 **4.3513** **<----- Critical F value**

Any obtained F value that is equal to or larger than 4.3513 will produce a rejection of the null hypothesis. Also, as a shortcut, recall that we can look at the p value and if that p value is = .05 or less, then we can reject the null hypothesis. In the present case, only the interaction term has a p value that is equal to or less than .05 and therefore it is the only null hypothesis that we can reject. The null hypothesis for the interaction is that the lines are parallel. Neither Factor A nor Factor B alone make any difference but, in combination, there is an interaction. What this means is that which level of A is best depends upon which level of B one is at.

Let's look quickly at a second example. Again, assume a two by two design where persons have been randomly assigned to one of the four cells. One independent variable is the instructional method and the second independent variable is the type of reinforcement given. The data look as follows.

		Level on Factor B		Means
		0	1	
Level on Factor A	0	13,16,14,11,9,10,8: **M=11.571**	12,10,9,9,11,8,6: **M=9.286**	**10.429**
	1	22,21,24,19,17,20,18: **M=20.143**	6,4,4,8,7,5,3: **M=5.286**	**12.714**
Means		**15.857**	**7.286**	**11.571**

A graph of the data is as follows.



Since the B0 line is higher than the B1 line, there probably is an effect due to Factor B. This can also be seen in the large difference in the column means (15.8 and 7.2). There also appears to be an "ordinal" pattern since the lines are not parallel and do not cross. We will wait to see if there is any impact due to Factor A since the difference in the row means (10.4 and 12.7) is not as large as for the column means. Here are the results from the analysis of variance.

ANALYSIS OF VARIANCE FOR THE PERFORMANCE DATA

Source	DF	SS	MS	F	p
Method	1	36.57	36.57	6.89	0.015
Reinf	1	514.29	514.29	96.86	0.000
Met*Rei	1	276.57	276.57	52.09	0.000
Error	24	127.43	5.31		
Total	27	954.86			

Using the p values as our guide, it appears that all of the F ratios are of a size that we should reject the null hypothesis in each case. We could have found the critical 95th percentile rank F value for 1 and 24 degrees of freedom also as a way to determine whether to retain or reject the null hypotheses. Any F ratio that is 3.0088 or larger will produce a rejection of the null hypothesis.

MTB > **invcdf .95;**
SUBC> **f 3 24.**
.9500 **3.0088**

Even though Factor B has an effect with B0 being better, the more interesting finding is the interaction that shows that the difference between the B0 and B1 lines is much greater at level A1 than at A0. It appears that the types of reinforcement do not make much of a difference for the first instructional method (Factor A level 0) but, makes a large difference if the second instructional method is used. In fact, if almost looks as if the first reinforcement method (B0) **elevates** performance whereas the second reinforcement method (B1) actually **depresses** performance. Overall, I think the message is to stay away from the second type of reinforcement!

Before leaving both the one factor and two factor analysis of variance designs, it should be mentioned that designs can be more complex than two factors. There can be 3 or more factors too! In addition, while I have concentrated on completely randomized designs, where people are randomly assigned to each cell in our design, other designs can have the same people operate under more than one condition. For example, we could have a design where there is an experimental and a control condition (random assignment to each of these) but then have each person provide a pretest and a posttest score. The pretest and posttest scores obviously are based on the same people. Factors where the same people perform at the different levels are sometimes called within subjects or repeated measures factors. In any case, these other types of designs call for a more complex analysis using ANOVA and we do not have space here to explore these.

Before moving on, it should be mentioned that ANOVA also has a set of assumptions for proper interpretation. These are very similar to those discussed in Chapter 18 on the "difference in means". Simply stated, people have to be randomly selected from the target populations and assigned randomly to the various conditions. Secondly, the populations have to be normally distributed with respect to the criterion measure. And last, the variances in the populations (on the criterion measure) also have to be equal. This last assumption (homogeneous variance assumption) is particularly important and violations of it can create serious misinterpretation problems.

Useful Minitab Commands

AOVO

Practice Problems

1. An investigator randomly assigns 10 people to each of 3 experimental groups in a study of drug dosage levels on reaction times (in minutes). The data are below. Make a graph of the cell means. What does the graph suggest? Then do a one factor ANOVA. Do you retain or reject the null hypothesis? If you reject the null hypothesis, do a Tukey follow-up test to better pinpoint the differences.

   Low Dosage: 12, 16, 14, 19, 17, 13, 17, 16, 14, 20

   Med Dosage: 19, 24, 17, 22, 19, 17, 16, 18, 21, 20

   High Dosage: 24, 19, 29, 26, 24, 27, 22, 26, 25, 29

2. In a study, a researcher manipulates both the number of times a student can read a passage and the number of self checking questions interspersed in the text. Eight students are randomly assigned to each of the 4 cells in the two by two design. After the study, a 20 item test over the material is given. The data are below. Make a graph of the cell means. What does that suggest? Then do the two factor ANOVA. Which null hypotheses do you retain and/or reject? Does either factor make a difference?

		B: # Questions Inserted		Means
		10	20	
A: # of Passage Readings	1	10,13,12,7,9,6,8,9	13,13,10,12,10,8,9,10	
	2	13,10,16,12,14,16,15,14	16,18,16,17,19,19,17,20	
Means				