

**Two-Factor Analysis of Variance (continued)
with solutions**

A plant pathologist was studying the effect of soybean cyst nematode infection on the activity level of a soybean gene suspected to play a role in plant defense against pathogen attacks. A total of 24 soybean plants were used in the experiment. Half of the plants were randomly selected to be inoculated with soybean cyst nematodes. The other half were mock inoculated with a control substance. The activity level of the gene was measured immediately after inoculation for 4 plants in each group. The activity level of the gene was measured 12 hours after inoculation for another 4 plants in each group. The activity level of the gene was measured 24 hours after inoculation for the 4 remaining plants in each group. The treatments are summarized below.

Treatment	Inoculation	Time Point	Number of Plants	Mean Activity Level
1	control	0	4	9.50
2	control	12	4	10.00
3	control	24	4	10.75
4	nematodes	0	4	9.75
5	nematodes	12	4	13.00
6	nematodes	24	4	17.25

1. This is an example of a two-factor experiment. Name the factors and the levels of each factor.

Factor A is Inoculation, there are 2 levels of this (control/nematodes)

Factor B is Time, there are 3 levels of this (0, 12, 24 hours)

So, this is a 2x3 factorial experiment. That is, there are 6 treatment groups, corresponding to level combinations of Inoculation and Time.

2. For two-factor experiments with two levels for each factor, we learned how to test for interaction between factors by testing whether a linear combination of treatment means was significantly different from zero. When one or more factors has more than two levels, we will test for interaction using the *F*-test provided in the *Type III Sum of Squares* portion of the SAS output. If there are *a* levels of one factor and *b* levels of the other factor, the *F*-test for interaction has numerator degrees of freedom $(a-1)(b-1)$ and denominator degrees of freedom $(n-ab)=(n-I)$ where *n* is the total sample size and *I* is the number of treatments. Note that the denominator degrees of freedom $(n-I)$ matches the degrees of freedom for error because the denominator of the *F*-statistic is *MSW*. Give the numerator and denominator degrees of freedom for the *F*-test for interaction in this example.

The general ANOVA table for testing for effects of Factor A (with *a* levels) and Factor B (with *b* levels) as well as interaction is given by

Source	DF	SS	MS	F
Factor A	$a-1$	Type III <i>SSA</i>	$MSA=SSA/(a-1)$	MSA/MSW
Factor B	$b-1$	Type III <i>SSB</i>	$MSB=SSB/(b-1)$	MSB/MSW
Interaction A*B	$(a-1)(b-1)$	Type III <i>SSAB</i>	$MSAB=SSAB/[(a-1)(b-1)]$	$MSAB/MSW$
Within Groups or Error	$n-I = n-ab$	<i>SSW</i>	$MSW=SSW/(n-ab)$	

You get the Type III Sum of Squares from SAS as well as *SSW*. Note: The Type III Sum of Squares and *SSW* do not necessarily have to add to the total: *SSTO*. These will when the data are "balanced," meaning when each of the $I=ab$ groups has the same number of observations. (In this case, Type I and Type III Sum of Squares will be equal and it is **always** true that Type I *SSA* + Type I *SSB* + Type I *SSAB* + *SSW* = *SSTO*. So when the data are balanced, the Type III Sum of Squares and *SSW* will sum to *SSTO*, just like Type I Sum of Squares.)

3. For two-factor experiments with two levels for each factor, we learned how to test for differences between the levels of one factor while averaging over the levels of the other factor by testing whether a linear combination of treatment means was significantly different from zero. This same strategy will work for any factor with only two levels, regardless of how many levels the other factor has. Test for a difference between control and treatment with nematodes by averaging over the levels of time in this example. To conduct the test, you need to know that $MSW=2.736$.

For Factor A (Inoculation), there are two levels: with and without nematodes. When there is no interaction, it makes sense to compare these two levels to check for differences. Whenever a factor has only two levels, you can compare these two levels using a t-test. This t-test for the two levels of Factor A will have to match the F-test based on the Type III Sum of Squares for Factor A (they are the same test).

Estimate of Overall Effect of Control Level = LSMEAN for control = $(9.50 + 10.00 + 10.75)/3$

Estimate of Overall Effect of Nematode Level = LSMEAN for nematode = $(9.75 + 13.00 + 17.25)/3$

Estimate of difference g = LSMEAN for nematode - LSMEAN for control = 3.25

$$SE(g) = \sqrt{MSW} \sqrt{\frac{(1/3)^2 + (1/3)^2 + (1/3)^2 + (-1/3)^2 + (-1/3)^2 + (-1/3)^2}{4}} = 0.6753$$

$$t = \frac{g - 0}{SE(g)} = 4.81 \quad df = n - I = n - ab = 24 - 6 = 18 \quad \text{two-sided p-value is less than 0.001}$$

strong evidence of a difference between control/nematode levels (nematode level has seemingly higher mean gene activity). Note the square of this "t" equals the F statistic from the Type III SS line for Inoculation (they give the same p-value though the p-value from the F statistics in SAS is the exact p-value; the p-value from the t-test above is just bounded. Again, you would test for overall differences in the levels of a factor A when there is no interaction between Factors A & B. If there is interaction, then you need to compare levels of factor A (control/nematode) separately over the levels of Factor B (i.e., for each level of time). In this case, there is evidence of interaction; see question 5 below.

4. When a factor has more than two levels, it is not possible to test for differences among levels of the factor by testing the significance of only one linear combination of means. Thus when a factor has more than two levels, we will test for the presence of any differences among the levels of the factor (averaging over the levels of the other factor) by using the F-test provided in the *Type III Sum of Squares* portion of the SAS output. If there are a levels of factor A and b levels of the other factor B, the F-test for differences among the levels of factor A has numerator degrees of freedom $(a-1)$, and the F-test for differences among the levels of factor B has numerator degrees of freedom $(b-1)$. Both tests have denominator degrees of freedom $(n-ab)=(n-I)$ because the denominator of the F-statistic is MSW . Compute treatment averages for each level of the time factor in this example, and give the numerator and denominator degrees of freedom for testing whether there are any significant differences among these averages.

Estimate of overall effect of time level 0 hrs: $(9.50 + 9.75)/2=9.625$

Estimate of overall effect of time level 12 hrs: $(10.00 + 13.00)/2=11.5$

Estimate of overall effect of time level 24 hrs: $(10.75 + 17.75)/2=14$

Note that we conduct a test for general differences among these 3 levels of time with the F-test from the Type III Sums of Squares which has $df1=b-1=3-1=2$. To test for differences among the levels of time, we can't again do the t-test (since there are 3 levels of time to compare); however, if you wished to do so, you could pick two levels of time and compare the estimates based on a t-test (this involves testing a linear combination of treatment group sample means, such as "estimate of 12 hr effect - estimate of 0 hr effect = $(1/2)(10.00)+(1/2)(13.00)+(-1/2)(9.50)+(-1/2)(17.75) = 11.5-9.625$ ") Again, the overall test for differences among time levels makes most sense when there are no interactions.

5. Examine the SAS code and output below. Is there evidence that this gene plays a role in defense against pathogen attack? Explain.

The F-test for interaction (see Type III SS for inoc*time) indicates strong evidence of interaction between the factors. You can see this in an interaction plot. The sample means for gene activity level are pretty close for the control/nematode at 0hrs; but the sample mean for activity in the presence of nematodes is much higher at 24 hrs compared to the control at 24 hrs. In other words, the differences in gene activity between control/nematode treatments is not the same over the factor ``time'' (this explains the interaction between inoculation and time).

Since there is significant interaction, we compare the control vs. nematode means at each level of time separately. In the SAS code below, you will find these comparisons and t-tests for differences.

```
proc glm;
  class inoc time;
  model y=inoc time inoc*time;
  lsmeans inoc time inoc*time;
  estimate 'nematode - control at time 0' inoc -1 1 inoc*time -1 0 0 1 0 0;
  estimate 'nematode - control at time 12' inoc -1 1 inoc*time 0 -1 0 0 1 0;
  estimate 'nematode - control at time 24' inoc -1 1 inoc*time 0 0 -1 0 0 1;
run;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
inoc	2	control nematode
time	3	0 12 24

Number of observations 24

Dependent Variable: y (activity level)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	179.7083333	35.9416667	13.14	<.0001
Error	18	49.2500000	2.7361111		
Corrected Total	23	228.9583333			

R-Square	Coeff Var	Root MSE	y Mean
0.784895	14.12771	1.654119	11.70833

Source	DF	Type I SS	Mean Square	F Value	Pr > F
inoc	1	63.37500000	63.37500000	23.16	0.0001
time	2	77.08333333	38.54166667	14.09	0.0002
inoc*time	2	39.25000000	19.62500000	7.17	0.0051

Source	DF	Type III SS	Mean Square	F Value	Pr > F
inoc	1	63.37500000	63.37500000	23.16	0.0001
time	2	77.08333333	38.54166667	14.09	0.0002
inoc*time	2	39.25000000	19.62500000	7.17	0.0051

Least Squares Means

inoc	y LSMEAN
control	10.0833333
nematode	13.3333333

time	y LSMEAN
0	9.6250000
12	11.5000000
24	14.0000000

inoc	time	y LSMEAN
control	0	9.5000000
control	12	10.0000000
control	24	10.7500000
nematode	0	9.7500000
nematode	12	13.0000000
nematode	24	17.2500000

Parameter	Estimate	Standard Error	t Value	Pr > t
nematode - control at time 0	0.25000000	1.16963907	0.21	0.8332
nematode - control at time 12	3.00000000	1.16963907	2.56	0.0195
nematode - control at time 24	6.50000000	1.16963907	5.56	<.0001