

5.1.1: Two-Factor Factorial - Greenhouse Example (SAS)

Let's return to the greenhouse example with *plant species* also as a predictive factor, in addition to *fertilizer type*. The study then becomes a 2×4 factorial as 2 types of plant species and 4 types of fertilizers are investigated. The total number of experimental units (plants) that are needed now is 48, as $r=6$ and there are 8 plant species and fertilizer type combinations.

The data might look like this:

		Fertilizer Treatment			
		Control	F1	F2	F3
Species	<u>A</u>	21.0	32.0	22.5	28.0
		19.5	30.5	26.0	27.5
		22.5	25.0	28.0	31.0
		21.5	27.5	27.0	29.5
		20.5	28.0	26.5	30.0
		21.0	28.6	25.2	29.2
	<u>B</u>	23.7	30.1	30.6	36.1
		23.8	28.9	31.1	36.6
		23.7	34.4	34.9	37.1
		22.8	32.7	30.1	36.8
		22.8	32.7	30.1	36.8
		24.4	32.7	25.5	37.1

The ANOVA table would now be constructed as follows:

Source	df	SS	MS	F
Fertilizer	$(4 - 1) = 3$			
Species	$(2 - 1) = 1$			
Fertilizer \times Species	$(2 - 1)(4 - 1) = 3$			
Error	$47 - 7 = 40$			
Total	$N - 1 = 47$			

The data presented in the table above are in unstacked format. One needs to convert this into a stacked format when attempting to use statistical software. The SAS code is as follows.

The data presented in the table above are in unstacked format. One needs to convert this into a stacked format when attempting to use statistical software. The SAS code is as follows.

```
data greenhouse_2way;
input fert $ species $ height;
datalines;
```

control	SppA	21.0
control	SppA	19.5
control	SppA	22.5
control	SppA	21.5
control	SppA	20.5
control	SppA	21.0
control	SppB	23.7
control	SppB	23.8
control	SppB	23.8
control	SppB	23.7
control	SppB	22.8
control	SppB	24.4
f1	SppA	32.0
f1	SppA	30.5
f1	SppA	25.0
f1	SppA	27.5
f1	SppA	28.0
f1	SppA	28.6
f1	SppB	30.1
f1	SppB	28.9
f1	SppB	30.9
f1	SppB	34.4
f1	SppB	32.7
f1	SppB	32.7
f2	SppA	22.5
f2	SppA	26.0
f2	SppA	28.0
f2	SppA	27.0
f2	SppA	26.5
f2	SppA	25.2
f2	SppB	30.6
f2	SppB	31.1
f2	SppB	28.1
f2	SppB	34.9
f2	SppB	30.1
f2	SppB	25.5
f3	SppA	28.0
f3	SppA	27.5
f3	SppA	31.0
f3	SppA	29.5
f3	SppA	30.0
f3	SppA	29.2
f3	SppB	36.1
f3	SppB	36.6
f3	SppB	38.7
f3	SppB	37.1
f3	SppB	36.8

```
f3      SppB      37.1
;
run;
/*The code to generate the boxplot
for distribution of height by species organized by fertilizer
in Figure 5.1*/

proc sort data=greenhouse_2way; by fert species;
proc boxplot data=greenhouse_2way;
plot height*species (fert);
run;
```

As a preliminary step in Exploratory Data Analysis (EDA), a side-by-side boxplot display of height vs. species organized by fertilizer type would be an ideal graphic. As the plot shows, the height differences between species are variable among fertilizer types (see for example the difference in height between *SppA* and *SppB* for *Control* is much less than that for *F3*). This indicates that *fert*species* could be a significant interaction prompting a factorial model with interaction.

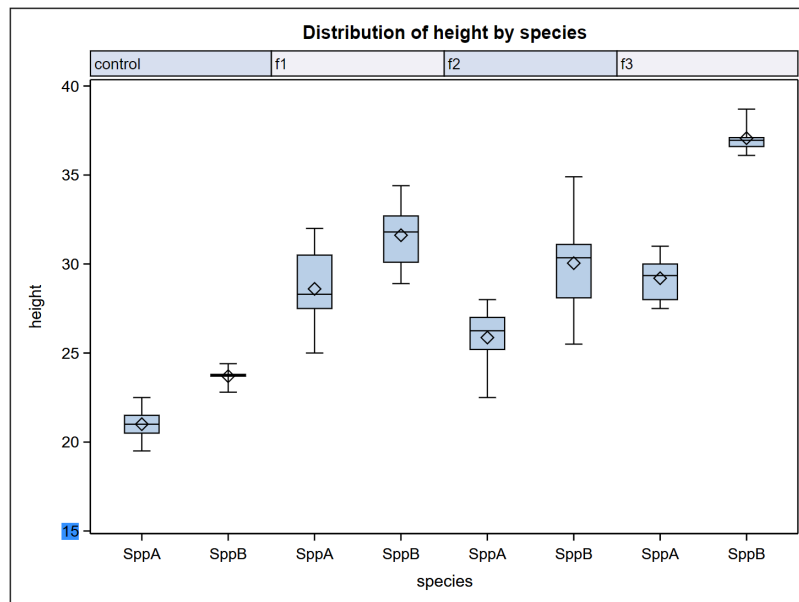


Figure 5.1.1.1: Boxplot for distribution of height by species organized by fertilizer.

To run the two-factor factorial model with interaction in SAS `proc mixed`, we can use:

```
/*Runs the two-factor factorial model with interaction*/
proc mixed data=greenhouse_2way method=type3;
class fert species;
model height = fert species fert*species;
store out2way;
run;
```

In the `proc mixed` procedure, similar to when running the single factor ANOVA. The name of the data set is specified in the `proc mixed` statement and so is the `method=type 3` option that specifies the way the F test is calculated. The `fert` and `species` factors that are both categorical are included in the class statement. The terms (or effects) in the model statement are consistent with the *source effects* in the layout of the "theoretical" ANOVA table illustrated in 5.1. Finally, the `store` command stores the elements necessary for the generation of the LS-Means interval plot.

Recall the two ANOVA rules, applicable to any model: (a). the df values add up to total df and (b). the sums of squares add up to total sums of squares. As seen by the output below, the df values and also the sums of squares follow these rules. (It is easy to confirm that the total sum of squares = 1168.732500, by the 2nd ANOVA rule.)

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
fert	3	745.437500	248.479167	Var(Residual)+Q(fert,fert*species)	MS(Residual)	40	73.10	<.0001
species	1	236.740833	236.740833	Var(Residual)+Q(species,fert*species)	MS(Residual)	40	69.65	<.0001
fert*species	3	50.584167	16.861389	Var(Residual)+Q(fert*species)	MS(Residual)	40	4.96	0.0051
Residual	40	135.970000	3.399250	Var(Residual)				

Rule

In a model with the interaction effect, the interaction term should be interpreted first. If the interaction effect is significant, then do NOT interpret the main effects individually. Instead, compare the mean response differences among the different factor level combinations.

In general, a significant interaction effect indicates that the impact of the levels of Factor A on the response depends upon the level of Factor B and vice versa. In other words, in the presence of a significant interaction, a stand-alone main effect is of no consequence. In the case where an interaction is not significant, the interaction term can be dropped and a model without the interaction should be run. See Section 5.1.1a: The Additive Model (No Interaction)).

Now applying the above rule for this example, the small p-value of 0.0051 displayed in the table above indicates that the interaction effect is significant, which means that the main effects of either *fert* or *species* should not be considered individually. It is the average response differences among the *fert* and *species* combinations that matter. In order to determine the statistically significant *fert* and *species* combinations, a suitable multiple comparison procedure, such as Tukey and Kramer procedure can be performed on the LS-Means of the interaction effect (i.e.: the treatment combinations).

The necessary follow-up SAS code to perform this procedure is given below.

```
ods graphics on;
proc plm restore=out2way;
lsmeans fert*species / adjust=tukey plot=(diffplot(center) meanplot(cl ascending)) cl
/* Because the 2-factor interaction is significant, we work with
the means for treatment combination*/
run;
```

SAS Output for the LSmeans:

fert*species Least Squares Means

fert*species Least Squares Means									
fert	species	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
control	SppA	21.0000	0.7527	40	27.90	<.0001	0.05	19.4788	22.5212
control	SppB	32.7000	0.7527	40	31.49	<.0001	0.05	22.1788	25.2212
f1	SppA	28.6000	0.7527	40	38.00	<.0001	0.05	27.0788	30.1212
f1	SppB	31.6167	0.7527	40	42.00	<.0001	0.05	30.0954	33.1379
f2	SppA	25.8667	0.7527	40	34.37	<.0001	0.05	24.3454	27.3879
f2	SppB	30.0500	0.7527	40	39.92	<.0001	0.05	28.5288	31.5712
f3	SppA	29.2000	0.7527	40	38.79	<.0001	0.05	27.6788	30.7212
f3	SppB	37.0667	0.7527	40	49.25	<.0001	0.05	35.5454	38.5879

Note that the p -values here ($Pr > t$) are testing the hypotheses that the fert and species combination means = 0. This may be of very little interest. However, a comparison of mean response values for different species and fertilizer combinations may prove to be more beneficial and can be derived from the diffogram shown in Figure 5.1.1.2 Again recall that, if the confidence interval does not contain zero, then the difference between the two associated means is statistically significant.

Notice also that we see a single value for the standard error based on the MSE from the ANOVA, rather than a separate standard error for each mean (as we would get from Proc Summary for the sample means). Again in this example, with equal sample sizes and no covariates, the *lsmeans* will be identical to the ordinary means displayed in the Summary Procedure.

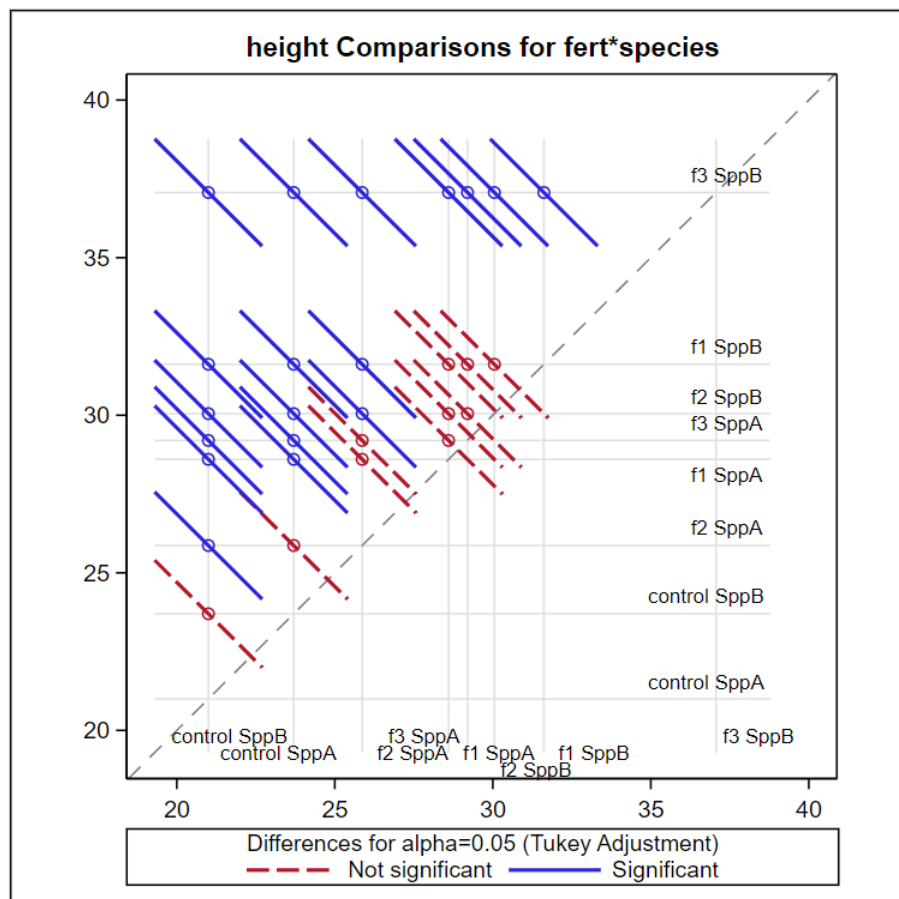


Figure 5.1.1.2: Diffogram for species and fertilizer combinations.

There are total of 8 *fert*species* combinations resulting a total of $\binom{8}{2} = 28$ pairwise comparisons. From the diffogram for differences in *fert*species* combinations, we see that 10 of them are not significant and 18 of them are significant at a 5% level after Tukey adjustment ([more about diffograms](#)). The information used to generate the diffogram is presented in the table for differences of *fert*species* least squares means in the SAS output (this table is not displayed here).

We can save the differences estimated in SAS `proc mixed` and utilize `proc sgplot` to create the plot of differences in mean response for the *fert*species* combinations as shown in Figure 5.1.1.3 The CIs shown are the Tukey adjusted CIs. SAS code to produce Figure 5.1.1.3 is not given in these notes. The interpretations of the plot are similar to what we observed from the diffogram in Figure 5.1.1.2

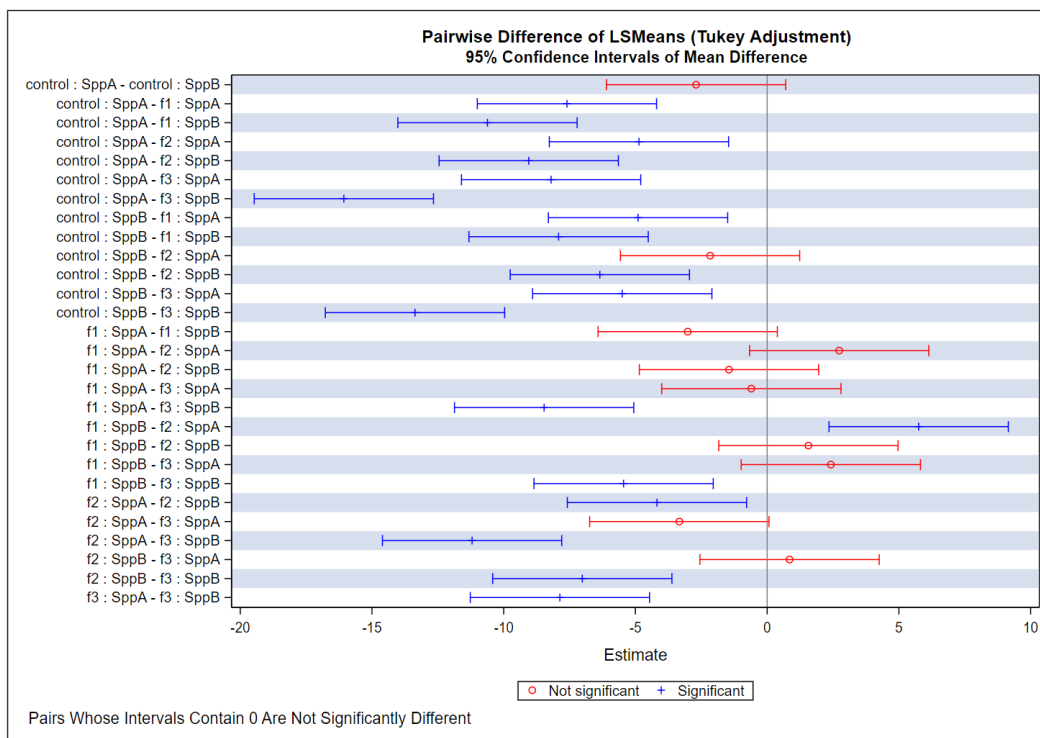


Figure 5.1.1.3: Plot of differences in mean response for the *fert*species* combinations.

In addition to comparing differences in mean responses for the *fert*species* combinations, the SAS code shared above will also produce the line plot for multiple comparisons of means for *fert*species* combinations (shown in Figure 5.1.1.4) and the plot of means responses organized in the ascending order with 95% CIs for *fert*species* combinations (shown in Figure 5.1.1.5).

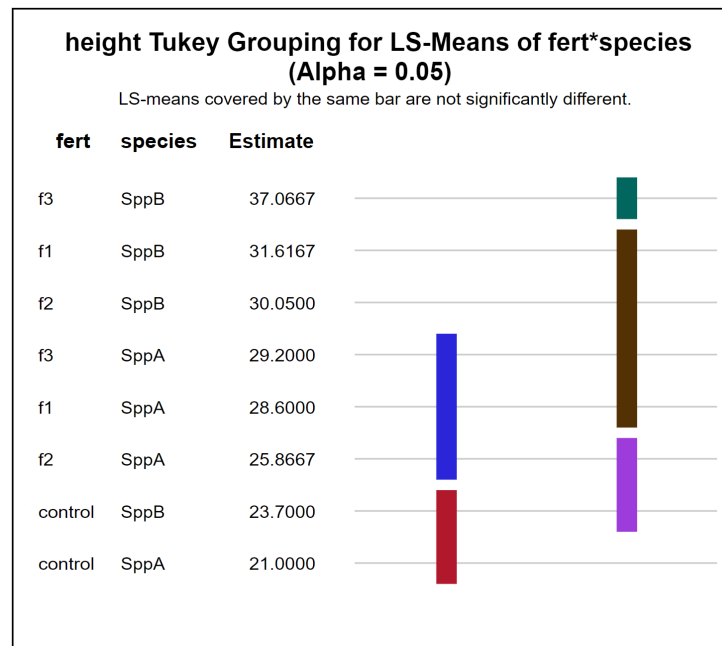


Figure 5.1.1.4: The line plot for multiple comparisons of means for *fert*species* combinations.

The line plot in Figure 5.1.1.4 connects groups in which the LS-means are not statistically different and displays a summary of which groups have *similar* means. The plot of means with 95% CIs in Figure 5.1.1.5 illustrates the same result, although it uses unadjusted CIs. We have organized the plot in the ascending order of estimated means to make it easy to draw conclusions.

Copy and Paste
Image here.
Delete this
placeholder image

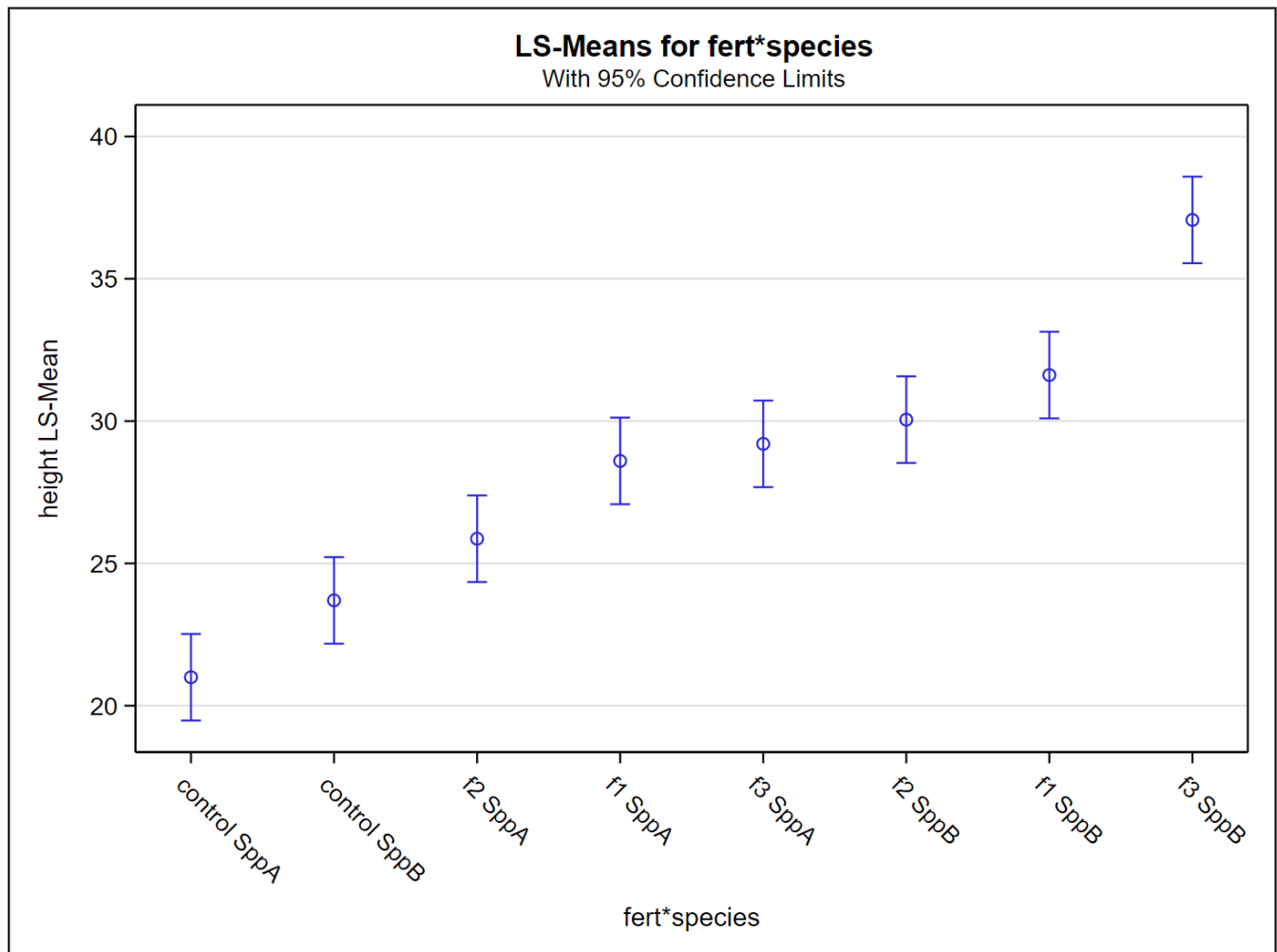


Figure 5.1.1.5: The plot of means with 95% CIs for *fert*species* combinations.

Using LSMEANS, subsequent to performing an ANOVA will help to identify the significantly different treatment level combinations. In other words, the ANOVA doesn't end with a *p*-value for an *F*-test. A small *p*-value signals the need for a mean comparison procedure.

This page titled [5.1.1: Two-Factor Factorial - Greenhouse Example \(SAS\)](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#).