

## A quick guide to writing contrast statements in SAS and SPSS

For fairly simple comparisons there are just a few steps to follow to get the proper coefficients to write statements in SAS and SPSS, but keeping some basic rules in mind will help for more complicated comparisons. (Stata does not have contrast functionality built-in but it can be added with some downloadable .ado files. Get them by typing “findit contrast” in Stata on a net-connected computer.)

1. Identify the order of factors and the way in which they are coded.

In SAS the factor ordering is specified in the class statement and in SPSS in the BY clause. *The order the variables appear in these places will dictate how you must write out the contrast.*

Also, the coding of variables is treated as either alphabetical or numeric by default. For example, if the variable sex has male and female are coded as M and F, the programs will expect a coefficient for *females* first.

2. It’s always a good idea to plot out your data and look at the means you are going to be comparing. SPSS profile plots are handy and easy on the eyes for this. In SAS a rough way of doing it is to run proc univariate on the dataset with all of the factors specified in a by statement. You’ll get a set of boxplots at the end of the output. Plots will help you figure out when you’ve coded something improperly and identify interesting comparisons you might not have thought of.

3. Identify the contrasts you are interested in.

For example, imagine a 3-way factorial CRD experiment with factors named A, B, and C. Factor A has 2 levels (1,2), factor B 3 levels (1,2,3) and factor C 4 levels (1,2,3,4).

*A simple comparison that can be planned is the difference between the levels of A when B is 2. How to find the coefficients for this will be shown below.*

4. Set up a table to hold the coefficients paying close attention to the order of factors and their levels. Factor A is first here and goes in the row; factor B goes on the column because of their placement in our class statement or by clause. Both have numerically ordered levels. The coefficients for the levels of our comparison are entered in the corresponding cells on the left and top margins of the table. They are shown in red italics here. These are the coefficients for A and B we will begin with. You should see where they come from because of the comparison being made ( $A_1$  vs.  $A_2$  in  $B_2$ ).

		B		
		1	2	3
A	1	<i>1</i>		
	2	<i>-1</i>		

5. The table is filled by *multiplying* the coefficients along the right and top and entering the value in the proper cell. The new values are shown in red here. These are the values for the A\*B interaction. They will be read into the contrast statement by going across each row starting with the top one. So A\*B would have 0 1 0 0 -1 0.

		B		
		1	2	3
		0	1	0
A	1	1	1	0
	2	-1	-1	0

6. Now *add* across the rows and down the columns to fill in the left and bottom rows of the table. These values will be used in the contrast statement for the main effects of A (on the right) and B (along the bottom). B has all zero coefficients that will not need to be written in the contrast statement. For A, we will have to write 1 -1 in that order.

		B		
		1	2	3
		0	1	0
A	1	1	1	0
	2	-1	-1	0
		0	0	0
		1	-1	0

7. Fill in the one remaining cell, which identifies a coefficient for the intercept, by *adding* across the bottom row and down the right column. No intercept will be specified because it is 0.

		B		
		1	2	3
		0	1	0
A	1	1	1	0
	2	-1	-1	0
		0	0	0
		1	-1	0

8. Finally the contrast/test/lmatrix statement can be written. For SAS glm or mixed one could write, estimate "A1 v A2 in B2" A 1 -1 A\*B 0 1 0 0 -1 0;  
 In SPSS glm one could write, /lmatrix "A1 v A2 in B2" A 1 -1 A\*B 0 1 0 0 -1 0.  
 The labels enclosed in the quotes are optional but help to identify what you're looking at in the output, particularly when multiple contrasts are being run.

NB. If you need to include values less than one as coefficients you can use decimal values in SAS going out at least 4 places (e.g. 0.3333). SPSS prefers fractions (e.g. 1/3).

Now imagine you have data from the same experiment but are interested in a more complicated interaction involving all three factors (A, B, and C). *Perhaps the last level of C (4) is a control, which you want to compare of the other levels to individually (1 v 4, 2 v 4, and 3 v 4), but you only are interested in when factor B is 2 and when A is 1.* We'll set up one of these individual contrasts below, comparing levels 1 and 4.

1. Begin the process by taking two of the factors for an initial table of coefficients. We'll take A and B.

		B			
		1	2	3	
		0	1	0	1
A	1	1	0	1	0
	2	0	0	0	0
		0	1	0	1

At this point it looks like both A and B remain in the contrast statement along with an intercept. But there is much more to be done.

2. Now check the A\*B interaction against C and find coefficients for the A\*B\*C interaction.

			C					
			1	2	3	4		
			1	0	0	-1		
A	1	B	0	0	0	0	0	
			2	1	1	0	-1	0
			3	0	0	0	0	0
	2	B	1	0	0	0	0	0
			2	0	0	0	0	0
			3	0	0	0	0	0
			1	0	0	-1	0	

Here we discover the intercept will drop out, as will the A\*B interaction. C will remain in the contrast statement.

3. We can continue to check if terms are need in the contrast by checking A and B against C which will remain now.

Get A\*C

			C				
			1	2	3	4	
			1	0	0	-1	
A	1	1	1	0	0	-1	0
	2	0	0	0	0	0	0
			1	0	0	-1	0

So A will drop out as will the intercept. C will remain.

Get B\*C

		C				
		1	2	3	4	
		1	0	0	-1	0
B	1	0	0	0	0	0
	2	1	1	0	-1	0
	3	0	0	0	0	0
		1	0	0	-1	0

So B will drop out along with the intercept (again), but C will remain.

4. Two final terms must be checked to find if they will enter the contrast statement. They are B\*C which can be checked against A, and A\*C which can be checked against B.

Do B\*C against A

				A		
				1	2	
				1	0	0
1	C	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
B	2	1	1	1	0	1
		2	0	0	0	0
		3	0	0	0	0
		4	-1	-1	0	-1
3	C	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
				0	0	0

A we knew was out, like the intercept, but B\*C will remain in.

Do A\*C against B

				B			
				1	2	3	
				0	1	0	1
1	C	1	1	0	1	0	1
		2	0	0	0	0	0
		3	0	0	0	0	0
		4	-1	0	-1	0	-1
A	2	1	0	0	0	0	0
		2	0	0	0	0	0
		3	0	0	0	0	0
		4	0	0	0	0	0
				0	0	0	0

B is out, along with the intercept, but A\*C will remain in.

5. The effects we've identified that will need to have coefficients specified are: C, A\*C, B\*C, and A\*B\*C. The contrast statement can then be written out as something like,

```
estimate "C1 v C4 in A1 and B2"
C 1 0 0 -1
A*C 1 0 0 -1 0 0 0 0
B*C 0 0 0 0 1 0 0 -1 0 0 0 0
A*B*C 0 0 0 0 1 0 0 -1 0 0 0 0 0 0 0 0 0 0 0 0 0;
```

or

/lmatrix "C1 v C4 in A1 and B2" followed by the same set of effects and coefficients.

NB. 0s that have no non-zero coefficients for an effect following them do not need to be written out. So A\*B\*C 0 0 0 0 1 0 0 -1 would have been equivalent to that line above.

6. The other contrasts we might be interested in comparing levels 2, and 3 to 4 can be written by shifting the 1 in the sequence of coefficients we have already identified.

For C 2 v 4 in A1 and B2

```
estimate "C2 v C4 in A1 and B2"
C 0 1 0 -1
A*C 0 1 0 -1 0 0 0 0
B*C 0 0 0 0 0 1 0 -1 0 0 0 0
A*B*C 0 0 0 0 0 1 0 -1 0 0 0 0 0 0 0 0 0 0 0 0 0;
```

and for C 3 V C4 in A1 and B2

```
estimate "C3 v C4 in A1 and B2"
C 0 0 1 -1
A*C 0 0 1 -1 0 0 0 0
B*C 0 0 0 0 0 0 1 -1 0 0 0 0
A*B*C 0 0 0 0 0 0 1 -1 0 0 0 0 0 0 0 0 0 0 0 0 0;
```

7. Your output can be "tagged" in SAS with an appropriate alpha for a desired experimentwise error rate. Because we'd be making three comparisons in this case we might have  $\alpha=.05$ . So an appropriate comparisonwise error rate would be  $.05/3=.0167$ . Adding `/alpha=.0167` to the end of the estimate statement will write this on each line of output pertaining to the contrast.