



Tutorials and worked examples for simulation,  
curve fitting, statistical analysis, and plotting.  
<http://www.simfit.org.uk>

Factorial ANOVA is designed to analyze the effects of categorical or quantitative variables, called factors, on observed quantities in order to detect the influence of these factors both separately and jointly.

### Example 1

Open the SIMFIT main menu, select[Statistics], [ANOVA], then [Factorial ANOVA], and analyze the data set below for no blocks but two factors, *A* and *B*, contained in the default SIMFIT test file `anova5.t.f1`.

Block	Factor A	Factor B	Observation
1	1	1	16.5
1	1	1	18.4
1	1	1	12.7
1	1	1	14.0
1	1	1	12.8
1	1	2	14.5
1	1	2	11.0
1	1	2	10.8
1	1	2	14.3
1	1	2	10.0
1	2	1	39.1
1	2	1	26.2
1	2	1	21.3
1	2	1	35.8
1	2	1	40.2
1	2	2	32.0
1	2	2	23.8
1	2	2	28.8
1	2	2	25.0
1	2	2	29.3

In order to perform factorial ANOVA using SIMFIT the data matrix must have a very precise structure which, in the case of the two factors above, is in standard order as follows.

1. Column 1 contains block numbers (in this case all 1 as there are no blocks)
2. Column 2 contains levels of factor *A* (in this case Level 1 = no hormone, Level 2 = hormone)
3. Column 3 contains levels of factor *B* (in this case Level 1 = female, level 2 = male)
4. Column 4 contains observed values (in this case of blood calcium in mg/100ml with 5 replicates)
5. Block numbers must be in nondecreasing order
6. Levels of *A* must be in nondecreasing order within each block
7. Levels of *B* must be in nondecreasing order within each level of *A*
8. There must be the same number of replicates in each group
9. Block numbers and factor levels must be consecutive integers  $\geq 1$

After analyzing the data in test file `anova5.tf1` SIMF<sub>IT</sub> displays the following results table.

Table 1: Results for Factorial ANOVA with test file `anova5.tf1`  
Transformation:  $x$  (untransformed data)

Source	<i>SSQ</i>	<i>NDOF</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Blocks	0.000	0	0.000	0.000	0.0000
Effect 1 ( <i>A</i> )	1386	1	1386	60.53	0.0000
Effect 2 ( <i>B</i> )	70.31	1	70.31	3.071	0.0989
Effect 3 ( <i>A * B</i> )	4.900	1	4.900	0.2140	0.6499
Residual	366.4	16	22.90		
Total	1828	19			

Treatment Means and Standard Errors

Overall mean	21.82
Treatment means	
Effect 1	13.50 30.15
Standard Error of difference in means:	2.140
Effect 2	23.70 19.95
Standard Error of difference in means:	2.14
Effect 3	14.88 12.12 32.52 27.78
Standard Error of difference in means:	3.026

Note that in factorial ANOVA tables the treatment effects are always output in standard order as follows.

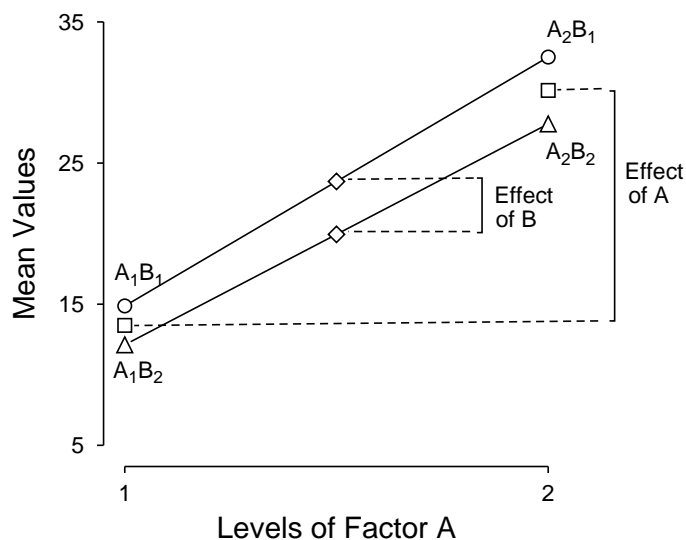
Effect 1:  $A_1, A_2$

Effect 2:  $B_1, B_2$

Effect 3:  $A_1B_1, A_1B_2, A_2B_1, A_2B_2$

Also, after analyzing the data, a plot like the next one can be useful to illustrate the effects of the factors.

### Means for Two-Factor ANOVA



### Example 2

To further emphasize the format required for factorial ANOVA, consider the following data set contained in SIMF<sub>IT</sub>test file `anova5.tf2`.

Block	Factor A	Factor B	Observation
1	1	2	361
1	1	3	253
1	2	1	325
1	2	2	317
1	2	3	339
1	3	1	326
1	3	2	402
1	3	3	336
1	4	1	379
1	4	2	345
1	4	3	361
1	5	1	352
1	5	2	334
1	5	3	318
1	6	1	339
1	6	2	393
1	6	3	358
2	1	1	350
2	1	2	340
2	1	3	203
2	2	1	397
2	2	2	356
2	2	3	298
2	3	1	382
2	3	2	376
2	3	3	355
2	4	1	418
2	4	2	387
2	4	3	379
2	5	1	432
2	5	2	339
2	5	3	293
2	6	1	322
2	6	2	417
2	6	3	342
3	1	1	82
3	1	2	297
3	1	3	133
3	2	1	306
3	2	2	352
3	2	3	361
3	3	1	220
3	3	2	333
3	3	3	270
3	4	1	388
3	4	2	379
3	4	3	274
3	5	1	336
3	5	2	307
3	5	3	266
3	6	1	389
3	6	2	333
3	6	3	353

This data set is for the yields of turnips in an agricultural experiment with 3 blocks, 6 levels of  $A$  (phosphate), and 3 levels of  $B$  (lime).

Table 2: Results for Factorial ANOVA with test5 file anova5.t.f2  
Transformation:  $x$  (untransformed data)

Source	$SSQ$	$NDOF$	$MS$	$F$	$p$
Blocks	30120	2	15060	7.685	0.0018
Effect 1 ( $A$ )	73010	5	14600	7.451	0.0001
Effect 2 ( $B$ )	21600	2	10800	5.510	0.0085
Effect 3 ( $A * B$ )	31190	10	3119	1.592	0.1513
Residual	66630	34	1960		
Total	222500	53			

Treatment Means and Standard Errors

Overall mean	331.1				
Block means	339.6	354.8	298.8		
Treatment means					
Effect 1	254.8	339.0	333.3	367.8	330.8
	360.7				
Standard Error of difference in means: 20.87					
Effect 2	334.3	353.8	305.1		
Standard Error of difference in means: 14.76					
Effect 3	235.3	332.7	196.3	342.7	341.7
	332.7	309.3	370.3	320.3	395.0
	370.3	338.0	373.3	326.7	292.3
	350.0	381.0	351.0		
Standard Error of difference in means: 36.14					

## Theory

The appropriate linear model for factorial ANOVA is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

where there are  $a$  levels of factor  $A$ ,  $b$  levels of factor  $B$  and  $n$  replicates per group, that is,  $n$  observations at each fixed pair of  $i$  and  $j$  values.

As usual,  $\mu$  is the mean,  $\alpha_i$  is the effect of  $A$  at level  $i$ ,  $\beta_j$  is the effect of  $B$  at level  $j$ ,  $\gamma_{ij}$  is the effect of the interaction between  $A$  and  $B$  at levels  $i$  and  $j$ , and  $e_{ijk}$  is the random error component at replicate  $k$ .

Also there are the necessary constraints on the parameters estimated, that is

$$\sum_{i=1}^a \hat{\alpha}_i = 0$$

$$\sum_{j=1}^b \hat{\beta}_j = 0$$

$$\sum_{i=1}^a \hat{\gamma}_{ij} = 0, \text{ for each } j, \text{ and}$$

$$\sum_{j=1}^b \hat{\gamma}_{ij} = 0 \text{ for each } i.$$

The null hypotheses would be

$$H_0 : \alpha_i = 0, \text{ for } i = 1, 2, \dots, a$$

to test for the effects of factor  $A$ ,

$$H_0 : \beta_j = 0, \text{ for } j = 1, 2, \dots, b$$

to test for the effects of factor  $B$ , and

$$H_0 : \gamma_{ij} = 0, \text{ for all } i, j$$

to test for possible  $AB$  interactions.

The analysis of variance table is based upon calculating  $F$  statistics as ratios of sums of squares that arise from the partitioning of the total corrected sum of squares as follows

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y} \dots)^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n [(\bar{y}_{i..} - \bar{y} \dots) + (\bar{y}_{.j.} - \bar{y} \dots) \\ &\quad + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots) + (y_{ijk} - \bar{y}_{ij.})]^2 \\ &= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y} \dots)^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y} \dots)^2 \\ &\quad + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ijk} - \bar{y}_{ij.})^2 \end{aligned}$$

It is clear from the  $F$  statistics and significance levels  $p$  in Table 1 table for `anova5.tf1` that, with these data,  $A$  has a large effect,  $B$  has a small effect, and there is no significant interaction. From table 2 for `anova5.tf2` the effects of blocking as well as  $A$  and  $B$  are significant, but the interaction between  $A$  and  $B$  is small.

Note that the factorial ANOVA table always outputs results in standard order, e.g.  $A_1B_1, A_1B_2, A_2B_1, A_2B_2$  and so on, while the actual coefficients  $\alpha_i, \beta_j, \gamma_{ij}$  in the model can be estimated by subtracting the grand mean from the corresponding treatment means. In the previous marginals plot, the line connecting the circles is for observations with  $B$  at level 1 and the line connecting the triangles is for observations with  $B$  at level 2. The squares are the overall means of observations with factor  $A$  at level 1 (13.5) and level 2 (30.15), while the diamonds are the overall means of observations with factor  $B$  (i.e. 23.7 and 19.95) from the results table. Parallel lines indicate the lack of interaction between factors  $A$  and  $B$  while the larger shift for variation in  $A$  as opposed to the much smaller effect of changes in levels of  $B$  merely reinforces the conclusions reached previously from the  $p$  values in the results table.

If the data set contains blocking, as with test files `anova5.tf2` and `anova5.tf4`, then there will be extra information in the ANOVA table corresponding to the blocks, e.g., to replace the values shown as zero in the results table for `anova5.tf1`, as there is no blocking with the data in `anova5.tf1`.

The SIMFIT factorial ANOVA test files illustrate the following examples of data formatting, and these can be consulted and subsequently analyzed if further clarification is required.

`anova5.tf1`: 0 blocks and 2 factors

`anova5.tf2`: 3 blocks and 2 factors

`anova5.tf3`: 0 blocks and 3 factors

`anova5.tf4`: 3 blocks and 3 factors