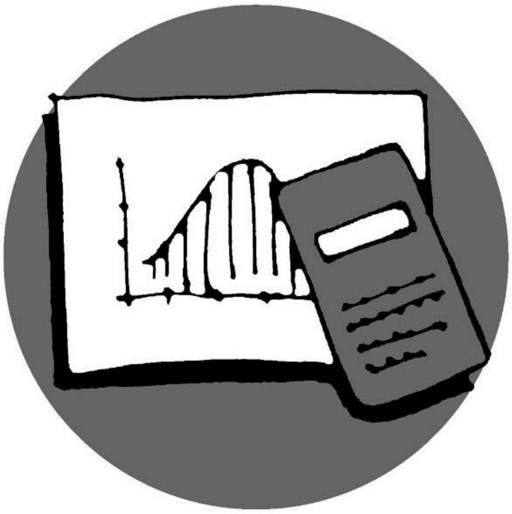
Library, Teaching and Learning

Factorial Designs

QMET201

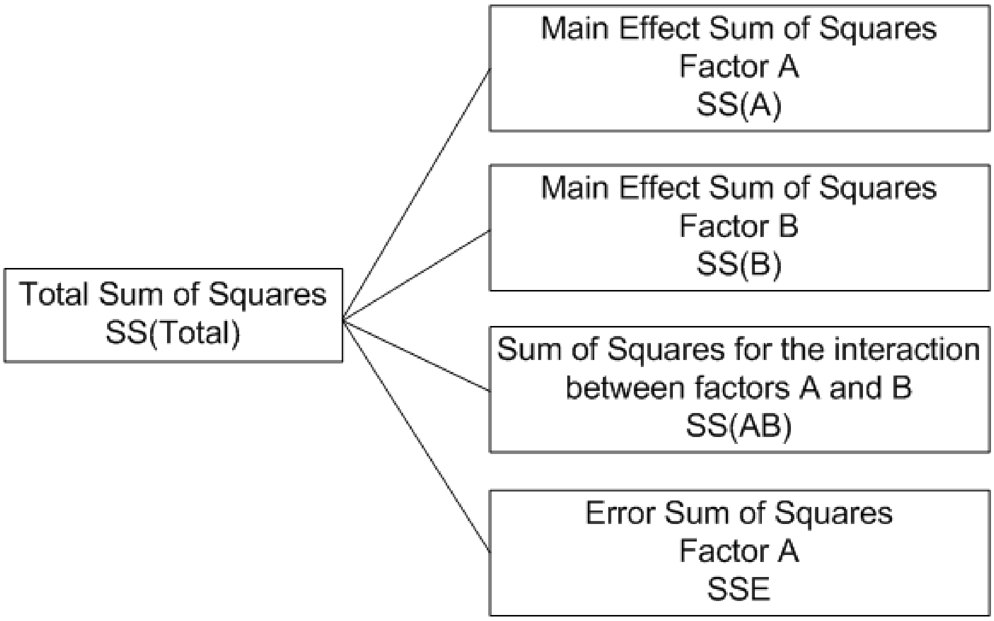


Lincoln University
Te Whare Wānaka o Aoraki
CHRISTCHURCH • NEW ZEALAND

New Zealand's specialist land-based university

Factorial Experiments
Analysis of variance for a *factorial experiment* allows investigation into the effect of two or more variables on the mean value of a response variable. Various combinations of factor 'levels' can be examined. It is normal to *replicate* a factorial experiment at least 2 times.

The analysis of variance for a factorial experiment is very similar to the previous ones. The sums of squares for the treatments are now replaced by sums of squares for the two factors, (called the main effect sums of squares) and the interaction sums of squares.



ANOVA TABLE

SOURCE	df	SS	MS	F	p
Blocks					
Factor A					
Factor B					
Interaction A × B					
Error					
Total					

$n = abr$ where a = number of levels of Factor A,

b = number of levels of Factor B,

r = number of replications of the factorial experiment.

ANOVA CALCULATIONS FOR A FACTORIAL EXPERIMENT

Calculate the degrees of freedom:

$$df_A = (a-1);\ df_B = (b-1);\ df_{A\times B} = (a-1)(b-1);$$

$$df_{total} = (abr-1);\ df_{error} = df_{total} - [df_A + df_B + df_{A\times B}] = ab(r-1)$$

Correction Factor:	$CF = \frac{(\sum all\ entries)^2}{number\ of\ entries} = \frac{(\sum y)^2}{n}$	
Total Sum of Squares:	$SS_{total} = \Sigma(each\ entry)^2 - CF = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$	
Main Effect Sum of squares: ***for Factor A*** ***for Factor B*** **…**	$SS_A = \frac{A_1^2 + A_2^2 + \ldots + A_a^2}{br} - CF$ $SS_B = \frac{B_1^2 + B_2^2 + \ldots + B_b^2}{ar} - CF$	Note: A_1, A_2 etc are *totals* for level 1, level 2 etc of Factor A. Also, the divisor for SS_A is $b \times r$ - the number of values included to make up each total, and for SS_B is $a \times r$. (Include for Factor C etc if necessary.)
Interaction Sum of squares:	$SS_{AB} =$ $\frac{AB_{11}^2 + AB_{12}^2 + \ldots + AB_{ij}^2}{r} - SS_A - SS_B - CF$	That is, square each cell total; add these totals; divide by the number of repeats (replications); then subtract SS_A , SS_B, and the correction factor.
Sum of Squares for Error:	$SS_{Error} = SS_{Total} - SS_A - SS_B - SS_{AB}$	
Mean Squares:	$MS_A = \frac{SS_A}{df_A}$, $MS_B = \frac{SS_B}{df_B}$, $MS_{AB} = \frac{SS_{AB}}{df_{A\times B}}$, $MS_{error} = \frac{SS_{error}}{df_{error}}$	
Test statistic *F*:	For testing main effect A: $\frac{MS_A}{MS_{error}}$ For testing main effect B: $\frac{MS_B}{MS_{error}}$ For testing AB interaction: $\frac{MS_{AB}}{MS_{error}}$	

If other factors are involved, these must be included, as well as the interactions.

For example, if there was a Factor C, $SS_C, SS_{AB}, SS_{AC}, SS_{BC},\ and\ SS_{ABC}$ would be required as well as the respective mean squares and F.

Note In the calculation of $LSD = t \times \sqrt{\frac{2 \times MSE}{r}}$, use df_{error} for t, and recalculate r for the specified LSD.

Worked examples

1. The concentration of the amino acid, alanine, was determined for four individuals of both sexes of three species of insects.

	Species A	Species B	Species C
Male	21.5	14.5	16.0
	19.6	17.4	20.3
	20.9	15.0	18.5
	22.8	17.8	19.3
Female	14.8	12.1	14.4
	15.6	11.4	14.7
	13.5	12.7	13.8
	16.4	14.5	12.0

$r = 4$

$a = 3\,(species)$

$b = 2\,(sexes)$

$n = 4 \times 2 \times 3 = 24$

The sums of the Sex by Species data are given as:

	spp			
Sex	1	2	3	All
1	84.8	64.7	74.1	223.6
2	60.3	50.7	54.9	165.9
All	145.1	115.4	129.0	389.5

Complete an ANOVA table:

Factor A = spp (3 levels)   Factor B = sex (2 levels); four replications

Source	DF	SS	MS	F	P
spp					
sex					
spp*sex					
Error					
Total					

- Fill in the degrees of freedom:
  $df_{total} = 3 \times 2 \times 4 - 1 = 23$; $df_{spp} = 2; df_{sex} = 1; df_{spp \times sex} = 2 \Rightarrow df_{error} = 23 - 5 = 18$

- Calculate Total Sum of squares: $SS_{total} = 21.5^2 + 14.5^2 + \ldots .14.5^2 + 12.0^2 - \frac{389.5^2}{24}$
  $= 238.89$
  (Suggest all values entered in calculator, then $SST = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$)

- Calculate spp, sex and spp × sex sums of squares:
  $$SS_{spp} = \frac{\left(145.1^2 + 115.4^2 + 129.0^2\right)}{8} - \frac{389.5^2}{24} = 55.261$$
  $$SS_{sex} = \frac{\left(223.6^2 + 165.9^2\right)}{12} - \frac{389.5^2}{24} = 138.72$$

$$SS_{(spp\times sex)} = \frac{\left(84.8^2 + 64.7^2 + 74.1^2 + 60.3^2 + 50.7^2 + 54.9^2\right)}{4} - SS_{spp} - SS_{sex} - \frac{389.5^2}{24}$$

$$= 6.89$$

- Calculate Error Sum of Squares

$$SS_{error} = SST - (all\ the\ rest) = 238.89 - \left(55.261 + 138.72 + 6.89\right) = 38.019$$

- Now fill in the ANOVA table:

Source	DF	SS	MS	F	P
spp	2	55.261	27.6305	13.08	<0.001
sex	1	138.720	138.720	65.68	0.000
spp*sex	2	6.89	3.445	1.63	*n.s.*
Error	18	38.018	2.112		
Total	23	238.89			

- Calculate a 1% LSD to test between Species means
The t value for two-tailed p=0.01 and 18 df is 2.8784. So

$$LSD = 2.8784 \times \sqrt{2.112 \times \frac{2}{8}} = 2.092$$

- Species 1 is found in boggy areas, whereas Species 2 & 3 are found in rather drier areas. Calculate the contrast sum of squares to test between Species 1 against the mean of Species 2 and 3.

This is just like the example worked through in class. First think of the mean of Spp 1 vs the mean of Spp2 and Spp3. The coefficients would be 1 and $-^1/_2$ *and* $-^1/_2$*. Multiply these by 2 to get 2 and -1, -1. Use these values in the following:*

$$ContrastSS = \frac{\left(2\times 145.1 - 1\times 115.4 - 1\times 129.0\right)^2}{\left(2^2 + (-1)^2 + (-1)^2\right)\times 8} = \frac{2097.64}{48} = 43.701$$

Pictorially, the experimental area might look like this:

Block 1

Nitrogen level 1 Moisture level 1	Nitrogen level 1 Moisture level 2	Nitrogen level 1 Moisture level 3	Nitrogen level 1 Moisture level 4
Nitrogen level 2 Moisture level 1	Nitrogen level 2 Moisture level 2	Nitrogen level 2 Moisture level 3	Nitrogen level 2 Moisture level 4
Nitrogen level 3 Moisture level 1	Nitrogen level 3 Moisture level 2	Nitrogen level 3 Moisture level 3	Nitrogen level 3 Moisture level 4

Block 2

Nitrogen level 1 Moisture level 1	Nitrogen level 1 Moisture level 2	Nitrogen level 1 Moisture level 3	Nitrogen level 1 Moisture level 4
Nitrogen level 2 Moisture level 1	Nitrogen level 2 Moisture level 2	Nitrogen level 2 Moisture level 3	Nitrogen level 2 Moisture level 4
Nitrogen level 3 Moisture level 1	Nitrogen level 3 Moisture level 2	Nitrogen level 3 Moisture level 3	Nitrogen level 3 Moisture level 4

2. Smith et al (1997, J. Appl. Ecol.) looked at the conservation value of grass pastures. Six blocks of grassland were marked out. Each block was divided into four plots and each plot randomly selected to receive one of the four treatment combinations from the following schematic table:

	Pasture Management	
	Conventional	Alternative
No fertilizer	14.52	34.74
Fertilizer	12.78	23.34

The error sum of squares was 6.0750.

i) Set out an analysis of variance table showing the degrees of freedom for the main treatment effects and the interaction, blocks, error and total.
ii) Calculate the sums of squares for main treatment effects and interaction.
iii) Calculate the mean squares for main treatment effects, interaction and error.
iv) Complete the analysis of variance table showing the F-ratios for treatment mean squares.
v) Put in the appropriate *p*-values from the F tables.
vi) Calculate the SED for the “Pasture management” means.

Solution:

First step is to complete the totals for the table:

	Pasture Management		
	Conventional	Alternative	Total
Fertilser	14.52	34.74	49.26
No Fertiliser	12.78	23.34	36.12
	27.3	58.08	85.38

A = pasture managements $\Rightarrow a = 2$, B = number of fertilizers $\Rightarrow b = 2$,
Number of replications: $r = 6 \Rightarrow n = 2 \times 2 \times 6 = 24$

Now complete the necessary calculations:

$$df_A = 1; df_B = 1; \; df_{A \times B} = 1 \times 1 = 1; df_{total} = (2 \times 2 \times 6 - 1) = 23; \; df_{error} = 23 - (1+1+1) = 20$$

$$CF = \frac{(\sum all\ entries)^2}{number\ of\ entries} = \frac{(\sum y)^2}{n} = \frac{85.38^2}{24} = 303.73925$$

$$Management\ SS = SS_A = \frac{A_1^2 + A_2^2 + \ldots + A_a^2}{br} - \frac{(\sum y)^2}{n} = \frac{27.3^2 + 58.08^2}{12} - CF = 39.4753$$

$$Fertiliser\ SS = SS_B = \frac{B_1^2 + B_2^2 + \ldots + B_b^2}{ar} - \frac{(\sum y)^2}{n} = \frac{49.26^2 + 36.12^2}{12} - CF = 7.19415$$

$$Interaction\ SS = SS_{AB} = \frac{AB_{11}^2 + AB_{12}^2 + \ldots + AB_{ij}^2}{r} - SS_A - SS_B - \frac{(\sum y)^2}{n}$$

$$= \frac{14.52^2 + 34.74^2 + 12.78^2 + 23.34^2}{6} - CF - Manage\ SS - Fert\ SS$$

$$= 3.88815$$

$$SS_{error} = 6.0750\,(given)$$

ANOVA table can now be completed:

Source	df	SS	MS	F	p
Block	5				
Management	1	39.47535	39.47535	97.47	<0.001
Fertiliser	1	7.19415	7.19415	17.763	<0.001
Manage*Fertiliser	1	3.88815	3.88815	9.600	<0.01
Error	15	6.0750	0.405		
Total	23				

(v) From tables, $F_{.05\ (1,\ 15)} = 4.54$, $F_{.01\ (1,\ 15)} = 18.68$, $F_{.001\ (1,\ 15)} = 16.59$

(vi) SED Management means = $\sqrt{\dfrac{EMS \times 2}{r}} = \sqrt{\dfrac{0.405 \times 2}{12}} = 0.2598$

3. The application of iron has sometimes been used to neutralise the toxic effect of metals like manganese (Mn), molybdenum (Mo) and vanadium (V) on legume crops. The following analysis of variance table is from an experiment measuring soybean growth. The experiment was a RCBD with five replicated blocks, and the treatments were a 3 by 4 factorial with iron (Fe) at three levels (2.5, 5.0, 20.0 ppm) and metal being "none", Mn, Mo, V. The measurement was log dry weight.

   Analysis of Variance for log_g

Source	DF	SS	MS	F	P
Block	4	0.00936	0.00234	0.45	0.773
Fe	*	*	*	148.88	0.000
Metal	3	3.1181	1.03727	198.47	0.000
Fe*Metal	6	0.74394	0.12399	*	*
Error	44	0.22996	0.00523		
Total	59	5.65122			

Here are the totals for blocks:

Block	Total
1	5.343
2	5.787
3	5.584
4	5.682
5	5.525
All	27.921

Here are the totals for the treatment combinations.

Rows: Fe Columns: Metal

	1	2	3	4	ALL
1	3.897	2.440	3.572	3.834	13.743
2	4.285	0.423	1.359	1.918	7.985
3	4.177	0.116	0.733	1.167	6.193
All	12.359	2.979	5.664	6.919	27.921

(i) Calculate the Sums of squares for Fe
(ii) Give the degrees of freedom for Fe
(iii) Calculate the mean square for Fe.
(iv) Determine the p value for the Fe*Metal interaction term.
(v) Calculate the 5%LSD to test for the difference between the 'Metal' means.

(vi) Explain how you would calculate the interaction sum of squares. (You are not required to calculate it, although you can if you want to.)

Solution (i) Enter Fe totals in your calculator. $SS_{Fe} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 1.55614$

(ii) There are 3 levels of Fe, $\Rightarrow$ there are 2 degrees of freedom.

(iii) Complete $MS_{Fe} = \frac{1.55614}{2} 0.778075$

(iv) $F_{interaction} = \frac{MS_{Fe*Metal}}{MS_{error}} = \frac{0.12399}{0.00523} = 23.707$

(v) Interaction degrees of freedom are 6,44 respectively. (For tables, use 40 df for the denominator df.) Table values are:

p	F
0.05	2.34
0.01	3.29
0.001	4.73

$F_{calc} > F_{tables}$.

That is, there is enough evidence to reject the null hypothesis.

4. Write out the ANOVA tables for EACH of the following three experiments. In each case give:
   a) the source of variance column
   b) the degrees of freedom column
   c) the 5% F value from the F table for each of the treatment and interaction lines.

i) ***Experiment A*** Six grass species, each replicated four times in a RCBD.

ii) ***Experiment B*** An RCBD with three blocks. The treatments are 3 soybean varieties, 4 potassium levels and 2 sulphur levels in a full factorial.

iii) ***Experiment C*** An experiment on corn. Six blocks were divided into two plots each and one randomly selected plot in each block allocated a high irrigation treatment; the other plot was not irrigated. Each plot was subdivided into 4 subplots and nitrogen at four levels applied.

Solution:

Experiment A $k = 6$; $r = 4$; $n = k \times r = 24$;

Source	df	5% F
Treatment (Grass)	5 (k-1)	2.90 (with $v_1, v_2 = 5, 15$)
Block	3 (r-1)	
Error	15	
Total	23 (n-1)	

Experiment B $a = 3$; $b = 4$; $c = 2$ $r = 3$; $n = a \times b \times c \times r = 3 \times 4 \times 2 \times 3 = 72$;

Source	df	5% F
Block	2	
Soybean	2	3.23
Potassium	3	2.84
Sulphur	1	4.08
Soybean* Potassium	6	2.34
Sulphur *Potassium	3	2.84
Sulphur* Soybean	2	3.23
Soybean* Sulphur *Potassium	6	2.34
Error	46	
Total	71	

Experiment C $r = 6,\ a = 2,\ b = 4,\ n = 6 \times 4 \times 2 = 48$

Source	df	5% F
Block	5	
Irrigation	1	4.17
Irrigation*Block	5	2.53
Nitrogen	3	2.92
Nitrogen*Irrigation	3	2.92
Error	30	
Total	47	

Advantages of two-way ANOVA

In one-way ANOVA, populations are classified according to one categorical variable, or factor. In the two-way ANOVA model, there are two factors, each with its own number of levels. When we are interested in the effects of two factors, a two-way design offers great advantages over several single factor studies:

- Valuable resources can be spent more efficiently, by studying two factors simultaneously rather than separately.
- The residual variation in the model can be reduced, by including a second factor thought to influence the response.
- We can investigate interactions between factors.

These considerations also apply to study designs with more than two factors. The choice of sampling or experimental design is fundamental to any statistical study. Factors and levels must be carefully selected by an individual or team who understands both the statistical models and the issues that the study will address.