

# 1

## An introduction to analysis of variance

---

### 1.4 The fertiliser dataset: an example of one-way ANOVA

In this first example, we wish to compare the efficacy of three fertilisers. A field experiment has produced the yield of a crop from 30 plots, each fertiliser having been applied to 10 of these plots. Which fertiliser produces the greatest yields?

#### ***Step 1: The data***

SPSS will require the data in columns as illustrated in Table 1.2; that is, with levels of fertiliser represented as subscripts 1 to 3 in one column called `FERTIL`, and the yields from each plot in a second column called `YIELD`. These text names must be entered at the top of each column, so that when you enter commands, these names can be used to refer to those columns.

#### ***Step 2: The question***

The relevant SPSS command is `ONEWAY` which stands for ‘oneway ANOVA’. To tell SPSS that `FERTIL` is a categorical variable, we use the key word ‘by’. (To indicate continuous variables the equivalent keyword is ‘with’). The menu route requires us to place `FERTIL` in the ‘Factors’ box. Factor is another word widely used in statistics for a categorical variable. When using the menu route, to transfer variable names between panes, you first need to highlight the variable name, then click on the arrow between the source and destination panes. SPSS will then move that variable to the destination pane.

SPSS COMMANDS 1.1 <b>Analysis of variance</b>	
Syntax	<pre>oneway YIELD by FERTIL       /statistics descriptive       /plot means.</pre>
Menu route	Analyze > Compare Means > One-way ANOVA YIELD → Dependent List FERTIL → Factor <div style="border: 1px solid black; display: inline-block; padding: 2px;">Options</div> <input checked="" type="checkbox"/> Descriptive <input checked="" type="checkbox"/> Means plot

Notice the structure of commands in SPSS. We typed the above command over three lines for clarity (and because we are limited to 80 characters per line) — but in fact SPSS ignores the carriage returns. Instead, it pays attention to the slash (/) which precedes each subcommand. The command then ends with a full stop. To execute these commands, the cursor must be placed within the command, or the command should be selected, before pressing the run button.

These commands should then give you the following output which falls into three parts. The first part is a table of sample sizes, means and standard deviations for each of the three treatment groups. SPSS also calculates the 95% confidence intervals for each group mean (as in Table 1.4 of the main text). The details of how these values are calculated from first principles are explained in the text. The second part is the ANOVA table. If you compare this to Box 1.1 of the main text, you will see that the essence of the table is the same. The column headed SIG contains the p-value. SPSS uses the phrase ‘between-groups’ to describe the sum of squares due to FERTIL, and ‘within-groups’ to describe the error sum of squares. Finally, SPSS plots the means, as requested.

SPSS OUTPUT BOX 1.1 **Analysis of the fertiliser data set: One-way ANOVA**

**One-way**

**Descriptives**

YIELD

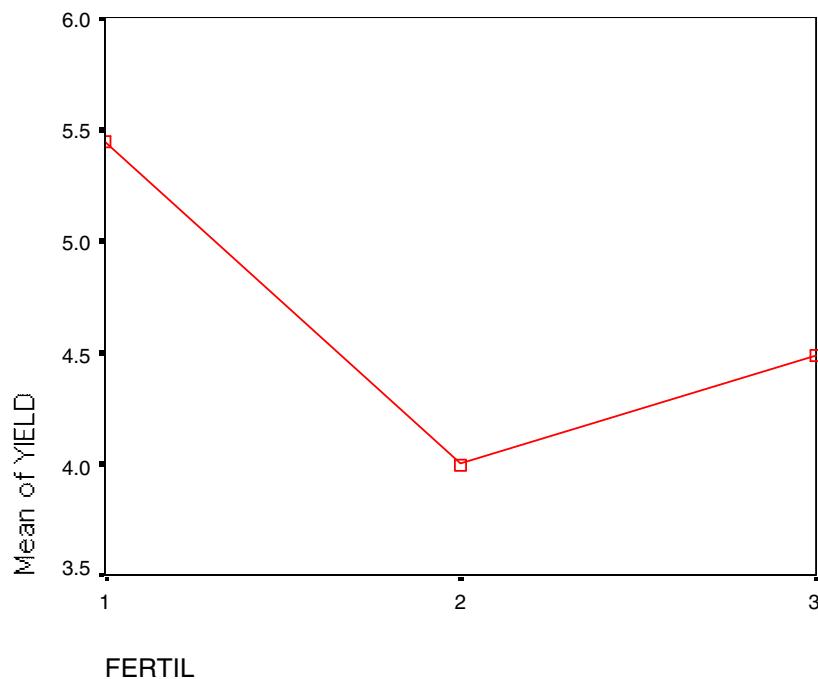
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	5.4450	.9760	.3086	4.7468	6.1432	4.07	7.14
2	10	3.9990	.9718	.3073	3.3039	4.6941	3.07	6.28
3	10	4.4870	.9747	.3082	3.7897	5.1843	3.53	7.00
Total	30	4.6437	1.1210	.2047	4.2251	5.0623	3.07	7.14

**ANOVA**

YIELD

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.823	2	5.411	5.702	.009
Within Groups	25.622	27	.949		
Total	36.445	29			

**Means plots**



8 An introduction to analysis of variance

Basic descriptive statistics may also be obtained, which are useful for having a preliminary inspection of the within group variances.

SPSS SYNTAX FOR TABLE 1.5 <b>Descriptive statistics for the <i>fertiliser</i> dataset</b>	
Syntax	means YIELD by FERTIL.
Menu route	Analyze > Compare Means > Means <div style="margin-left: 40px;">YIELD → Dependent List</div> <div style="margin-left: 40px;">FERTIL → Independent List</div>

SPSS OUTPUT FOR TABLE 1.5 <b>Descriptive statistics for the <i>fertiliser</i> dataset</b>						
<b>Means</b>						
<b>Case Processing Summary</b>						
	Cases					
	Included		Exclude		Total	
	N	Percent	N	Percent	N	Percent
YIELD * FERTIL	30	60.0%	20	40.0%	50	100.0%
<b>Report</b>						
YIELD						
FERTIL	Mean	N	Std. Deviation			
1	5.4450	10	.97598			
2	3.9990	10	.97175			
3	4.4870	10	.97471			
Total	4.6437	30	1.12104			

---

## 1.7 Exercises

### Melons

This experiment compares the yield from four melon varieties. It was designed so that each variety was grown in six plots—but two plots growing variety 3 were accidentally destroyed. There is just one categorical explanatory variable. We have analysed it here using two separate sets of commands as follows:

SPSS COMMANDS FOR TABLE 1.7 <b>ANOVA for melons</b>	
Syntax	<code>oneway YIELDM by VARIETY .</code> <code>means YIELDM by VARIETY .</code>
Menu route	Analyze > Compare Means > One-Way ANOVA YIELDM → Dependent List VARIETY → Factor Analyze > Compare Means > Means YIELDM → Dependent List VARIETY → Independent List

Which would produce the following output:

SPSS OUTPUT FOR TABLE 1.7 <b>Analysis of variance for melons</b>						
<b>One-way</b>						
<b>ANOVA</b>						
YIELDM						
	Sum of Squares	df	Mean Square	F	Sig.	
Between Groups	1115.281	3	371.760	23.79	.000	
Within Groups	281.189	18	15.622			
Total	1396.470	21				
<b>Means</b>						
<b>Case Processing Summary</b>						
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
YIELDM * VARIETY	22	44.0%	28	56.0%	50	100.0%
<b>Report</b>						
YIELDM						
VARIETY	Mean	N	Std. Deviation			
1	20.4900	6	4.69442			
2	37.4033	6	3.95050			
3	20.4625	4	4.76469			
4	29.8967	6	2.22989			
Total	27.6632	22	8.15466			

**Dioecious trees**

See SPSS output for this exercise in the answers for exercises.