

GenStat ANOVA workshop 1

Contents

- | | |
|-----------------------------|---|
| 1. Introduction | 7. Comparing means using LSDs |
| 2. A problem to analyse | 8. Saving means and residuals |
| 3. Performing the ANOVA | 9. Alternative ways of specifying the model |
| 4. Checking assumptions | 10. Further information |
| 5. Interpreting the output | 11. Reference |
| 6. Plotting graphs of means | 12. Acknowledgement |

Course pre-requisites: Attendees must have attended a GenStat Introduction workshop (or be familiar with GenStat); and must understand the use of ANOVA, blocking, and factorial designs.

1. Introduction

GenStat does Analysis of Variance (ANOVA) very well, and can be used to analyse complicated experimental designs relatively easily.

GenStat differs from most other statistics packages in that with GenStat you specify the treatment factors and blocking factors separately for an ANOVA. GenStat then arranges the factors appropriately in the ANOVA table. This is especially useful with more complicated nested designs such as split-plot which may have two or more error terms.

This workshop will not analyse complicated designs, but will instead concentrate on the simpler randomised block, 2-way factorial design.

2. A Problem to Analyse

The data being analysed is from Clewer & Scarisbrick (2001) (Example 12.3, p.166). A factorial trial was carried out to determine the response of a variety of oilseed rape to 3 levels of plant growth regulator (PGR) and 5 levels of nitrogen fertiliser (N). This is a 3x5 factorial design and was replicated in 3 randomised blocks. The 15 treatment combinations (Table 1) were randomly allocated to plots within each block (Table 2).

Table 1 Treatment combinations.

Treatment	PGR	N	Treatment	PGR	N	Treatment	PGR	N
1	1	1	6	2	1	11	3	1
2	1	2	7	2	2	12	3	2
3	1	3	8	2	3	13	3	3
4	1	4	9	2	4	14	3	4
5	1	5	10	2	5	15	3	5

Table 2 One possible layout for randomizing the 15 treatment combinations within the blocks.

Block 1					Block 2					Block 3				
7	2	6	11	9	3	14	5	13	2	14	3	9	7	13
8	3	12	5	4	6	8	9	11	1	4	5	12	1	11
13	15	10	1	14	12	10	4	15	7	15	10	2	6	8

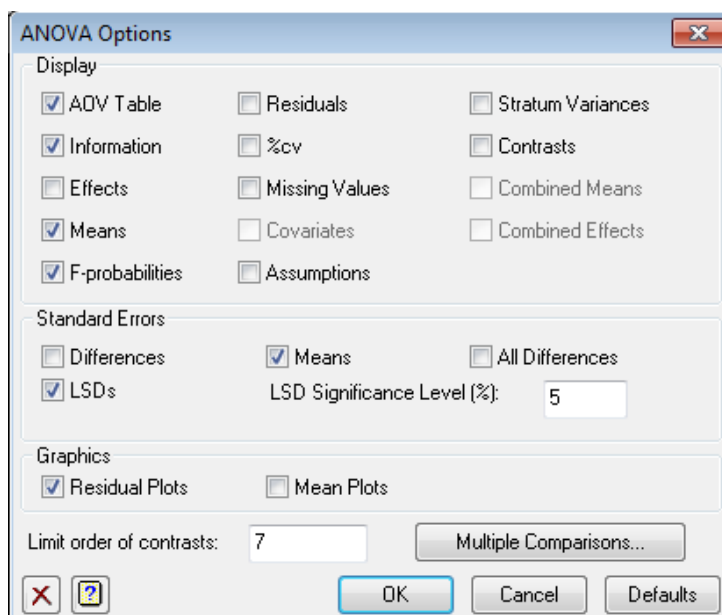
3. Performing the ANOVA

First start GenStat, then load the GenStat data file (`Oilseed rape yield.gsh`).

The data file contains 45 rows of data, with values for `yield` (a variate), and `Block`, `Nitrogen`, and `PGR` (the factors).

- Use Stats > Analysis of Variance > General...
- Select Design: Two-way ANOVA (in Randomized Blocks)
- Y-Variate: `yield`
- Treatment 1: `Nitrogen` Treatment 2: `PGR`
(Note: it doesn't matter which factor is Treatment 1 and which is Treatment 2.)
- Blocks: `Block`
- Click [Options...] then under 'Standard Errors' uncheck the Differences box and instead check the boxes for Means and LSDs.

Also check the box for Residual Plots (under 'Graphics') - these plots will enable you to check the assumptions necessary for the ANOVA.



- Click [OK] to return to the main dialog box (Figure 1).
- Click [Run] to produce the ANOVA and the Residual plots.

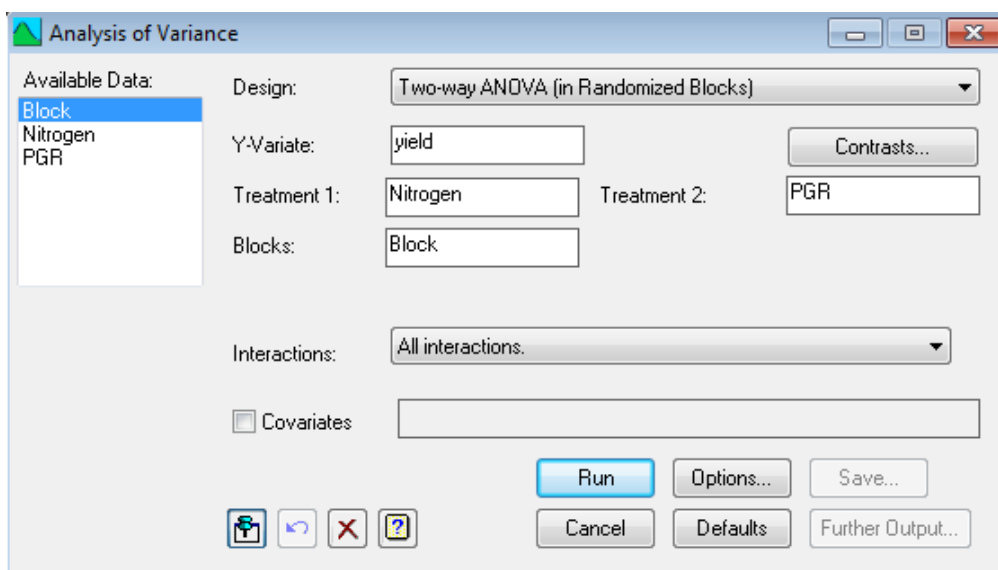


Figure 1 ANOVA dialog box for two-way ANOVA with randomized blocks.

4. Checking the assumptions

Residual plots provide a way of checking the normality and the equal variance assumptions of the ANOVA.

By default, four types of residual plot will be produced (Figure 2). Graphs A, C, and D (Figure 2) can be used to see if the residuals are normally distributed. I prefer to use C. If the graph forms a straight line through the origin (0, 0), the residuals are perfectly normally distributed. In this case, the graph shows a more or less straight line which indicates that the residuals are near enough to normally distributed. The s-bend at the extremes of the graph indicates the residuals are somewhat bunched at the tails but in this example the line is sufficiently straight for us to accept the assumption of normality.

Graph B shows that the variances are about equal regardless of the fitted value, and can be considered homogeneous. This is an important assumption, because although the ANOVA procedure is robust to departures from normality (so that you can still analyse somewhat non-normal data) it is less robust against departures from equal variances. If the variances are not equal, then the F values and P-values may be incorrect and give misleading conclusions. For this analysis, the assumptions of the ANOVA appear to be met and we can continue to interpret the output.

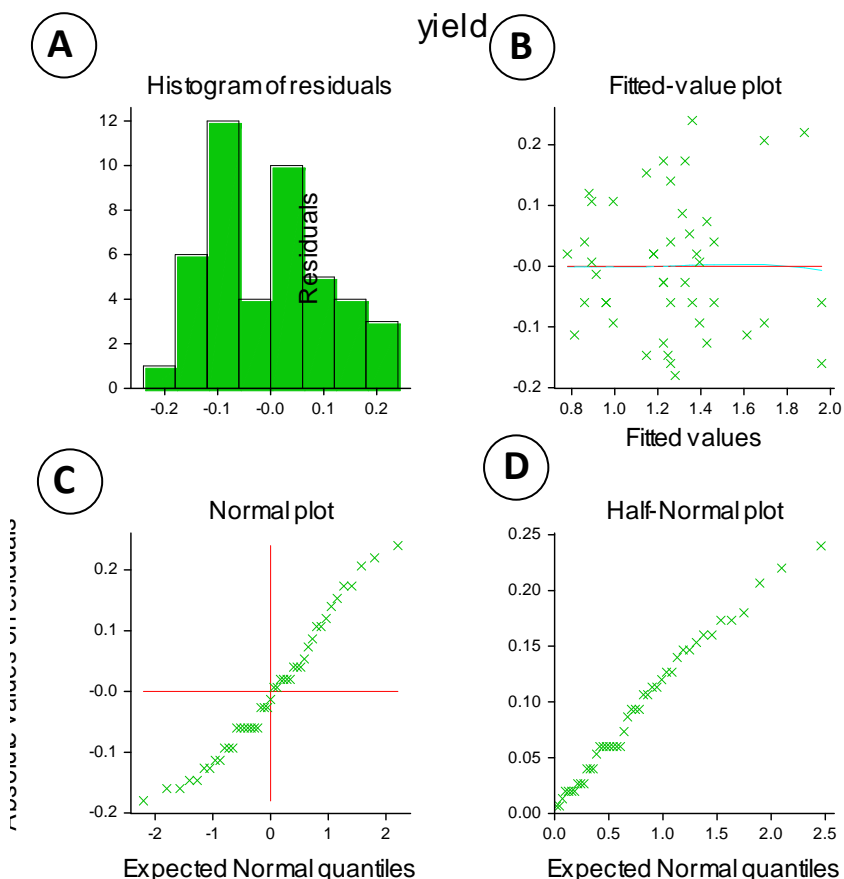


Figure 2 Residual plots produced after ANOVA. (A) Histogram of residuals, (B) Residuals vs Fitted values, (C) Normal plot of residuals, (D) Half-normal plot of residuals.

5. Interpreting the output

GenStat divides the ANOVA table into strata. The Block stratum can generally be ignored as we are not interested in difference between blocks, and GenStat does not provide a P-value for Block.

In the Block.*Units* stratum we can see that there is a significant interaction ($P < 0.001$) as well as significant main effects of Nitrogen ($P < 0.001$) and PGR ($P = 0.023$). Note that GenStat uses the headings v.r. (variance ratio) instead of F-value, and F pr. (F probability) instead of P-value.

```

Analysis of variance
=====
Variate: yield

Source of variation      d.f.      s.s.      m.s.      v.r.      F pr.
Block stratum           2         0.06400   0.03200   1.63
Block.*Units* stratum
Nitrogen                 4         2.49022   0.62256   31.73   <.001
PGR                      2         0.16933   0.08467   4.32    0.023
Nitrogen.PGR            8         1.01511   0.12689   6.47   <.001
Residual                 28        0.54933   0.01962
Total                    44        4.28800
    
```

By default, GenStat reports any unusual data values that are more than two standard deviations away from the fitted value (which can be thought of as a treatment mean adjusted for block differences).

You should always check any values with large residuals to see if there is a reason for them, e.g. the data were entered incorrectly, or an error was made when measuring that particular sample.

```
* MESSAGE: the following units have large residuals.
Block 1      *units* 9          0.240   s.e. 0.110
```

The means are listed for all of the main effects and the interaction.

```
Tables of means
=====
Variate: yield
Grand mean  1.260
Nitrogen    1      2      3      4      5
           0.878  1.133  1.356  1.567  1.367
PGR         1      2      3
           1.347  1.220  1.213
Nitrogen    PGR      1      2      3
  1          0.933  0.833  0.867
  2          1.233  0.967  1.200
  3          1.400  1.300  1.367
  4          1.933  1.333  1.433
  5          1.233  1.667  1.200
```

The standard error of the mean (e.s.e value) is given for each factor and for the interaction means. These S.E.s are often displayed as error bars on graphs of the means, and can be used to calculate confidence intervals.

```
Standard errors of means
-----
Table          Nitrogen          PGR          Nitrogen
                PGR
rep.            9            15            3
d.f.            28            28            28
e.s.e.          0.0467         0.0362         0.0809
```

The 5% LSD (Least Significant Difference) values are given and can be used to compare means among the treatments. Usually, the LSD is not used unless the ANOVA shows there is a significant difference.

```
Least significant differences of means (5% level)
-----
Table          Nitrogen          PGR          Nitrogen
                PGR
rep.            9            15            3
d.f.            28            28            28
l.s.d.          0.1353         0.1048         0.2343
```

6. Plotting graphs of means

It may be useful to see a graph of the treatment means.

- After performing the ANOVA, we can get graphs of means by clicking [Further Output...], then [Means Plots...]. Complete the 'Means Plot' dialog box as shown below (Figure 3).

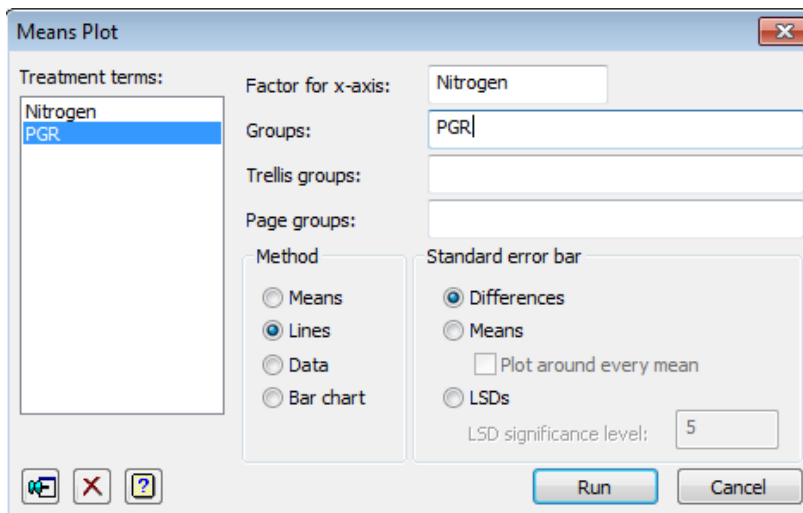


Figure 3 Dialog box to produce an interaction plot of treatment means.

- Click [Run] to produce the graph (Figure 4). The graph appears in a window in the GenStat Graphics Viewer.

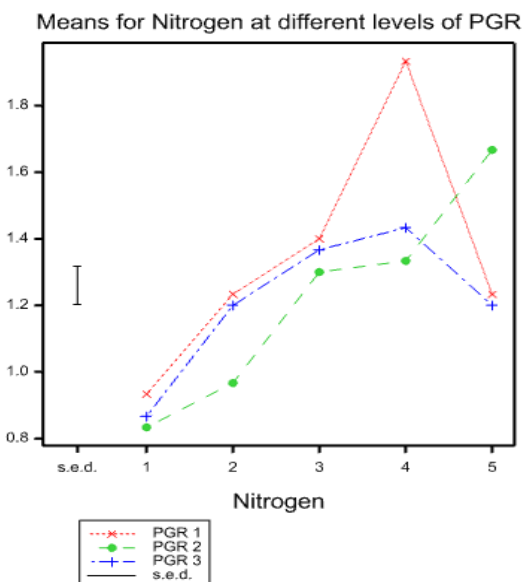


Figure 4 Graph of Nitrogen and PGR interaction means. The vertical bar represents the standard error of the difference between means.

- You can edit the graph by clicking Edit, Edit Graph... in the menu bar at the top of the Graphics Viewer.

Experiment with changing the title, adding axis labels and changing the default symbols and line colour/style/method. (Figure 5)

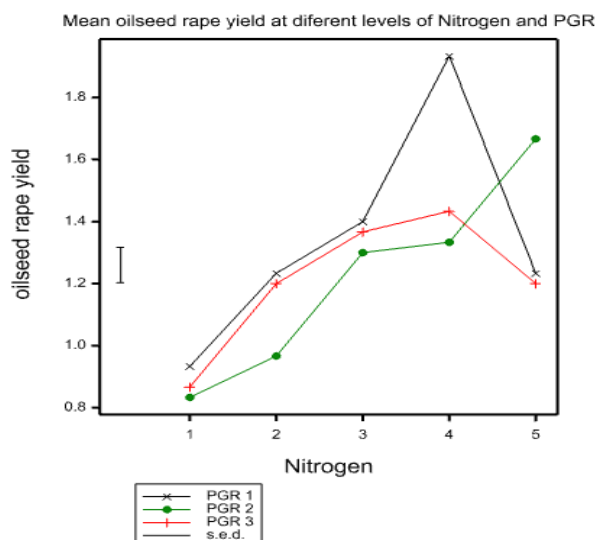


Figure 5 Graph of Nitrogen and PGR interaction means edited to change the title, add axis labels and to change the default symbols. Also note that the label for the sed has been removed from the x axis.

- The 'Means Plot' dialog box also allows you to plot standard error of mean values alongside the graph (or plotted around every mean). You can also plot LSD values as error bars.

Using the 'ANOVA Further Output' dialog box, access the 'Means Plot' dialog box again and experiment with plots of standard error of mean values or LSD values.

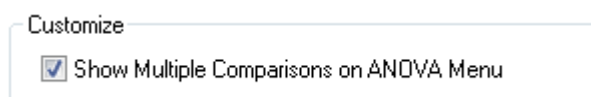
7. Comparing means using LSDs

Once you have determined that there is a significant difference between means (using the ANOVA output), it may be useful to determine just which means are significantly different from each other. There are a number of multiple comparison methods available. One commonly used method is Fisher's Protected Least Significant Difference, or LSD, for short.

Automatic multiple comparisons

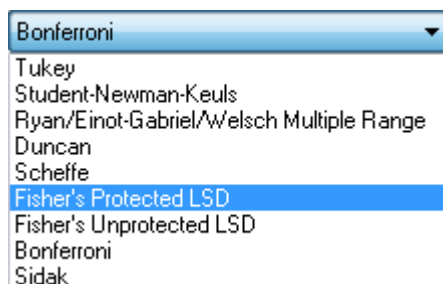
GenStat has a limited facility to automatically compare means. The menu option for this is not active by default – you first have to turn the Multiple Comparisons menu on. To do this, use:

- Tools > Options... > then on the 'Menus' tab, check the 'Show Multiple Comparisons' box:



We will compare the five Nitrogen means using LSD. To do this:

- Return to the ANOVA dialog box and click [Further Output...], then [Multiple Comparisons...]
- Select Treatment: Nitrogen
- Check the box beside 'Multiple Comparisons' and then select 'Fisher's Protected LSD' from the drop down list:



- To display both 'Means with letters' and 'Means with lines' check both of these boxes in the 'Display' section.
- Click [OK] (to complete the 'Multiple Comparisons' dialog)
- Click [Run] (to complete the 'Further Output' dialog) and the following Multiple Comparisons output should appear in the output window:


```

Fisher's protected least significant difference test
=====

Nitrogen
-----

          Mean
1         0.878
2         1.133
3         1.356 |
5         1.367 |
4         1.567

          Mean
1         0.878 a
2         1.133 b
3         1.356 c
5         1.367 c
4         1.567 d

```

- Repeat 'Fisher's Protected LSD' test, leaving the 'Means with lines' and 'Means with letters' boxes checked but now also checking the 'Description' and 'Comparisons' boxes. Notice the additional output produced by these.
- Also note that you can present the means in either Ascending (default setting) or Descending order (under 'Sort Means').

Understanding the Multiple Comparisons output

The output from 'Means with letters' shows the mean for each level of Nitrogen with letters alongside. Means with letters in common are 'not significantly different'.

In the output from 'Means with lines', means with a line in the same horizontal position are 'not significantly different'.

To help you understand the multiple comparison process for allocating lines and letters, the remainder of this section describes how to compare pairs of means in a methodical way *by hand*.

To carry out the Fisher's Protected LSD process by hand, you first need to calculate (or locate from the output), the appropriate LSD value (least significant difference) and list the means in numerical order. Add the LSD value to the first (smallest) mean to get a 'sum'. Any of the other means that are smaller than this sum are not significantly different from the first mean. Record this by drawing a line from the first mean across all of the non-significant means. Repeat the process by now adding the LSD to the 2nd smallest mean, and comparing this second sum to the larger means. If any of the larger means are smaller than this second sum, then those means are not significantly different to the 2nd mean. Again, draw a line across these non-significant means but only if the line will extend further than the first line (if it doesn't, then don't draw a line). This process is repeated until non-significance is reached with the last (largest) mean.

For example, to compare the five Nitrogen means with each other (Table 3):

- The Nitrogen 5% LSD = 0.1353. Adding the smallest mean and the LSD gives $0.878 + 0.1353 = 1.0133$. In this case, no other means are less than 1.0133. This means that the lowest mean is significantly different from all of the other treatment means.
- Adding the LSD to the 2nd mean gives $1.133 + 0.1353 = 1.2683$. Again, none of the higher means are less than this sum. So, the 2nd mean is significantly different to the higher means.
- Adding the LSD to the 3rd mean gives $1.356 + 0.1353 = 1.4913$. In this case, the 4th mean (for Nitrogen 5) is less than this sum so that the means for Nitrogen 3 and 5 are not significantly different from each other. (Note the two vertical lines beside these means in the 'Means with lines' output.)
- Lastly, adding the LSD to the 4th mean gives $1.367 + 0.1353 = 1.5023$. This sum is lower than the 5th (and last) mean (for Nitrogen 4), so the 4th and 5th means are significantly different from each other.

A completed table of working is shown in Table 4. The usual presentation method is to replace the lines with lowercase letters, and to reorder the means into a meaningful treatment order (Table 5).

Table 3 Mean oilseed rape yield under different levels of Nitrogen. Note that the means are shown in the order lowest mean value to highest mean value.

Nitrogen	Mean
1	0.878
2	1.133
3	1.356
5	1.367
4	1.567

Table 4 Table of means showing lines from working. Means joined with lines are not significantly different at $P = 0.05$.

Nitrogen	Mean
1	0.878
2	1.133
3	1.356
5	1.367
4	1.567

Table 5 Table of mean oilseed rape yield under different levels of nitrogen. Means with the same letter are not significantly different at $P = 0.05$.

Nitrogen	Mean	
1	0.878	a
2	1.133	b
3	1.356	c
4	1.567	d
5	1.367	c

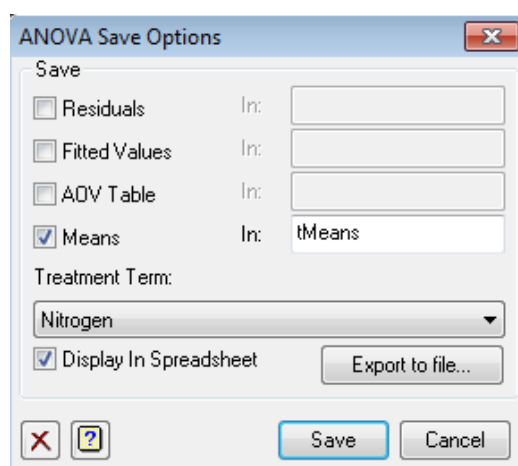
The comparison between means using the LSDs could also be carried out for the PGR means and for the interaction means using the appropriate LSD values.

8. Saving means and residuals

To save the values of the means, you can either copy and paste the values from the Output window, or save the means in a table. Each method has advantages: copying and pasting is quick and easy; saving to a table allows storing of means with more decimal places of accuracy and the saving of the values to a spreadsheet or other file.

To save the means to a table and display in a GenStat spreadsheet:

- Return to the 'ANOVA' dialog box
- Click [Save...]
- Complete the 'ANOVA Save Options' dialog box as follows – note the means will be saved in a GenStat *table* structure named 'tMeans' and will be displayed in their own GenStat spreadsheet.



- To save the results in an Excel spreadsheet, return to the 'ANOVA Save Options' dialog box and click [Export to file...]
- Check the boxes to specify the results you wish to save and click [Browse...] to select the location where you want to store the .xlsx file.
- Click [Save] to close the dialog box and create the Excel file.
- Open the newly created Excel file - depending on your 'Save options' choices, there may be a number of separate worksheets in the Excel file (see the tabs at the bottom of the spreadsheet window).

9. Alternative ways of specifying the model

There are many designs to choose from in the Analysis of Variance dialog box and often the menu will list the appropriate design for your experiment. However, sometimes you will need to select the 'General Analysis of Variance' design and specify the model yourself by using model operators.

The operators used to construct models in GenStat are:

- The summation operator (+) - used to add terms together.
For example, for factors A and B, the model $A+B$ would give the main effects of A and B.
- The dot operator (.) - used to specify an interaction.
For example, the interaction between A and B is given by $A.B$
- The crossing operator (*) - used as a shortcut to specify main effects and interactions.
For example, $A*B = A + B + A.B$
- The nesting operator (/) - used to specify terms that are nested. It is most often used with blocks to specify a split-plot design, e.g., $Blocks/MainPlots/SubPlots$.

There are various ways to specify the design for the two-way ANOVA with randomised blocks that we have been analysing:

- 1) Treatment Structure: Nitrogen*PGR
Block Structure: Block
- 2) Treatment Structure: Nitrogen + PGR + Nitrogen.PGR
Block Structure: Block
- 3) Treatment Structure: Block+Nitrogen*PGR
Block Structure: *(leave empty)*

Note that not all statistical analysis software packages use the same operators as GenStat. For example, in Minitab, interaction is specified by * and the shortcut operator for specifying main effects and interactions is the | operator.

10. Further information

- GenStat's *Anova and Design Guide* (available via Help > GenStat Guides > Anova and Design...) describes the analysis of a number of experimental designs.

This guide also includes a number of practical exercises using data files that are available within GenStat (File > Open Example Data Sets...).

- For revision of ANOVA (and other biometrical methods), you may find the following texts useful (all held in the Lincoln University Library):

Clewer, A. G., & Scarisbrick, D. H. (2001). *Practical statistics and experimental design for plant and crop science*. Chichester ; New York : Wiley.

Harraway, J. (1997). *Introductory statistical methods for biological, health and social sciences*. [Dunedin, N.Z.] : University of Otago Press.

Ott, R.L., & Longnecker, M. (2010). *An introduction to statistical methods and data analysis* (6th ed.). Belmont, CA : Brooks/Cole Cengage Learning.

Zar, J.H. (1999). *Biostatistical analysis* (4th ed.). Upper Saddle River, N.J, Prentice Hall.

- You'll find a range of statistics lessons using GenStat at the following University of Otago web page:

<http://www.maths.otago.ac.nz/video/statistics/GenStatLessons/>

11. Reference

Clewer, A. G., & Scarisbrick, D. H. (2001). *Practical statistics and experimental design for plant and crop science*. Chichester ; New York : Wiley.

12. Acknowledgements

Andrew McLachlan (previously of Lincoln University) designed a number of workshops for versions of GenStat available prior to 2007. The workshop presented today is based around one of the original workshops, updated for use with GenStat 16th edition (©2013).

Alison Lister
Learning Advisor (Mathematics and Statistics)
Library, Teaching and Learning