Solution to Exercise 11.4 (Version 1, 3/5/15)

from Statistical Methods in Biology: Design & Analysis of Experiments and Regression (2014) S.J. Welham, S.A. Gezan, S.J. Clark & A. Mead. Chapman & Hall/CRC Press, Boca Raton, Florida. ISBN: 978-1-4398-0878-8

Exercise 11.4

In Exercises 6.3 and 8.5 you analysed the score of potato scab from a CRD. Repeat your analysis and plot the residuals in field layout (as defined by the factors Row and Col). Identify any clear spatial trend and add suitable extraneous factors into the model to account for this. Compare your new model with the original, and comment on whether the increased complexity is justified.

Solution 11.4

The factor allocation is shown in Table S11.4.1.

Table S11.4.1. Frequency table for factor combinations in potato scab trial.

		Sulphur dose			
		0	300	600	1200
Timing of application	None	8	-	-	-
	Autumn	-	4	4	4
	Spring	-	4	4	4

The structure has all factorial combinations of sulphur dose and autumn/spring application plus added control treatments. We can describe this using a nested explanatory structure (see Section 8.5). There is no structural component as the experiment was set up as a CRD. In Exercise 6.3 we established that a logit transformation gave acceptable residual plots. The full model can be written as

Response variable: *logit*
Explanatory model: Type/(Sulphur*Timing)

where *logit* = logit(*Scab*) and factor Type is defined to have level 1 for plots with the control treatment and level 2 otherwise and factors Sulphur and Timing were defined in the data file. Residuals from this model are plotted against row number and column numbers in Figure S11.4.1. There is no obvious pattern of residuals against row number, but a consistent pattern in residuals across columns: residuals in columns 5, 6 and 7 are consistently larger than residuals in columns 1 and 2.

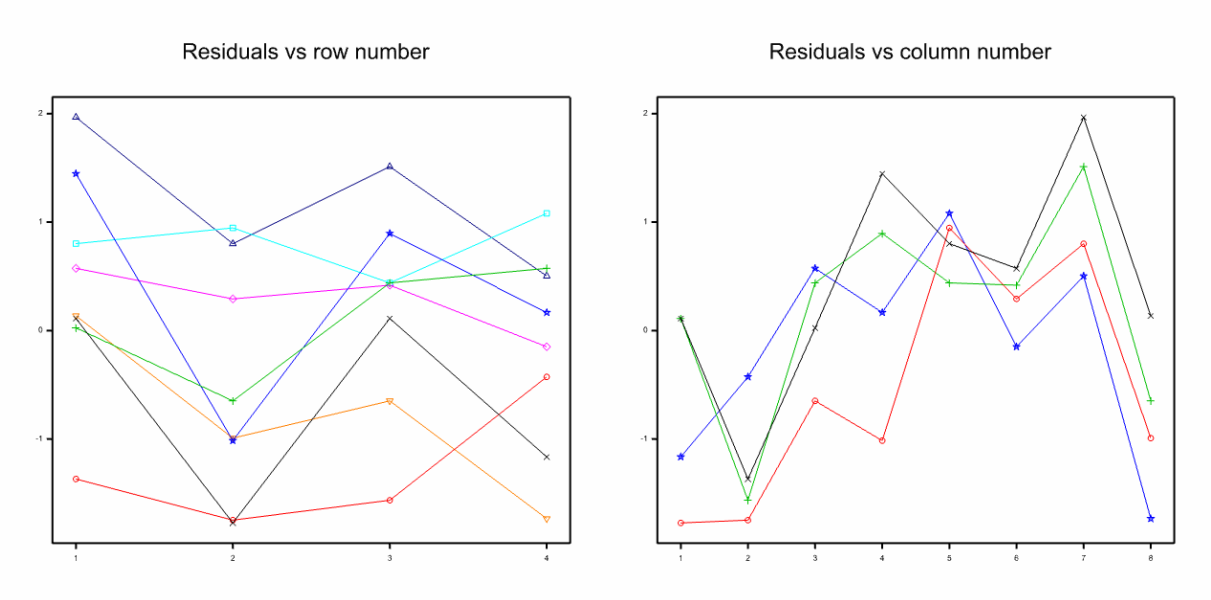

Figure S11.4.1. Residuals from logit(*Scab*) plotted against row number grouped by column (left) or column number grouped by row (right).

To make sure this pattern does not affect the results, we can adjust for it, by fitting the Col factor as an extraneous term. Ideally we would fit this as a structural term in the model, but as it was not included in the design it is not balanced with respect to the treatment terms and so cannot then be analysed by multi-stratum ANOVA. We therefore use the intra-block analysis with model

Response variable: *logit*
Explanatory model: Col + Type/(Sulphur*Timing)

The ANOVA tables with and without the extraneous term are shown in Table S11.4.2. Diagnostic plots of the residuals show no cause for concern, and plotting the residuals from this analysis against rows and columns show that the pattern has been removed (Figure S11.4.2).

Table S.11.4.2 Sequential ANOVA tables for potato scab data without and with extraneous column factor. Ty denotes factor Type, S denotes factor Sulphur and Ti denotes factor Timing. SS = sum of squares, MS = mean square, VR = variance ratio.

Sequence 1				
Source	df	SS	MS	VR
+ Type	1	3.0983	3.0983	10.601
+ Ty.S	2	2.2142	1.1071	3.788
+ Ty.Ti	1	2.1897	2.1897	7.492
+ Ty.S.Ti	2	0.6828	0.3414	1.168
Residual	25	7.3067	0.2923	
Total	31	15.4916		

Sequence 2				
Source	df	SS	MS	VR
+ Col	7	6.9657	0.9951	7.245
+ Type	1	1.9758	1.9758	14.385
+ Ty.S	2	1.6753	0.8376	6.099
+ Ty.Ti	1	1.5676	1.5676	11.413
+ Ty.S.Ti	2	0.8350	0.4175	3.040
Residual	18	2.4723	0.1373	
Total	31	15.4916		

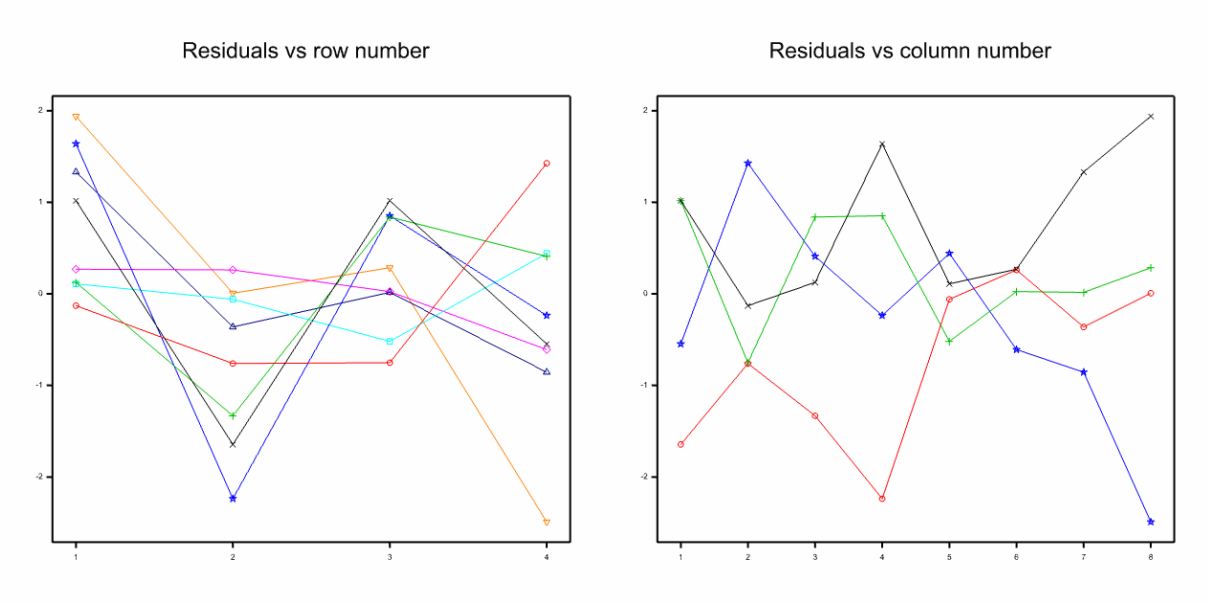

Figure S11.4.2. Residuals from logit(*Scab*) model with extraneous column factor (Col) plotted against row number grouped by column (left) or column number grouped by row (right).

The extraneous column term uses 7 df, reducing the residual df from 25 to 18 df, which still leaves sufficient df to give a reasonable estimate of the residual variance. The estimated residual variance is reduced substantially from 0.292 to 0.137 (the residual mean square) and the variance ratio for the Col term ($F_{7,18}$ = 7.245, $P < 0.001$) gives very strong evidence of column effects that cannot be accounted for by background variation. We conclude that removing the column effects has accounted for some extraneous variation. However, as columns were not orthogonal to the treatment effects, removing the column term has also reduced some of the treatment sums of squares (although not the Type.Sulphur.Timing interaction). However, this effect is outweighed by the reduction in the residual mean square and so the variance ratios for the treatment terms increase. The interpretation of the ANOVA table does not change substantially: the Type.Sulphur.Timing interaction is still not statistically significant ($F_{2,18}$ = 3.040, P = 0.073), and all other terms have an increased observed significance level (smaller P) but again no real change in inference.

Can we justify using the extraneous factor here? In this case, probably yes, as we have clear evidence from the residual plots of a spatial pattern in a biological scenario (field trials) where this is often observed and real, and we have sufficient residual df that the power of the analysis is not greatly compromised (we could check this using the methods of Chapter 10). In general we might be cautious if the residual df were small, if the spatial pattern was biologically unlikely or implausible or if there was little impact of fitting the extraneous factor. We should also be mindful of the danger of over-fitting where our evidence for extraneous factors comes from examining the data rather than from knowledge of events that occurred in the experiment (eg. the pigeon grazing observed in Example 15.4).