Solution to Exercise 17.6 (Version 1, 11/10/15)

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 17.6

An experiment was done to measure the response of yield to dose of nitrogen fertilizer. The design was a RCBD with four blocks of five treatments, corresponding to 0, 50, 100, 150 and 200 kg/ha of nitrogen applied and the response is plot yield. File FERTILIZER.DAT contains the unit numbers (*ID*), structural factors (Block, Plot), the amount of nitrogen applied (variate *N*) and the plot yields (variate *Yield*). Find a non-linear model to describe the response of yield to applied nitrogen. Check your model for misspecification and lack of fit. Write down and interpret the predictive model.

Data 17.6 (FERTILIZER.DAT)

ID	Block	Plot	N	Yield	ID	Block	Plot	N	Yield
1	1	1	50	17.4	11	3	1	50	17.0
2	1	2	200	22.1	12	3	2	100	20.3
3	1	3	100	18.1	13	3	3	200	21.2
4	1	4	150	21.3	14	3	4	150	23.2
5	1	5	0	11.5	15	3	5	0	11.2
6	2	1	150	20.5	16	4	1	100	18.5
7	2	2	200	20.5	17	4	2	150	21.9
8	2	3	0	8.5	18	4	3	200	22.4
9	2	4	100	17.4	19	4	4	50	14.8
10	2	5	50	13.5	20	4	5	0	12.0

Solution 17.6

Figure S17.6.1 shows the crop yields plotted against fertilizer application for each block. Systematic block differences are visible. The response to fertilizer dose shows a sharp increase at lower doses that tails off at higher doses. This matches an exponential curve with a negative slope parameter (β), and so this is the model we will fit. This model can be written as

$$y_{ij} = \alpha + \beta \exp(-\gamma x_{ij}) + e_{ij} ,$$

where y_{ij} is the yield from the j^{th} plot in the i^{th} block with fertilizer dose x_{ij} and model deviation e_{ij}. As stated above, we expect $\beta < 0$ for the shape in Figure S17.6.1. In principle, we would always like to include the structural component of the model, and especially here where we can see block differences in Figure S17.6.1. The standard assumption behind a RCBD design is that each block might have a different mean, but treatment differences are similar across blocks. We can build this assumption into our non-linear model by allowing different intercepts of the exponential curve for each block, ie. by fitting a separate parameter α for each block.

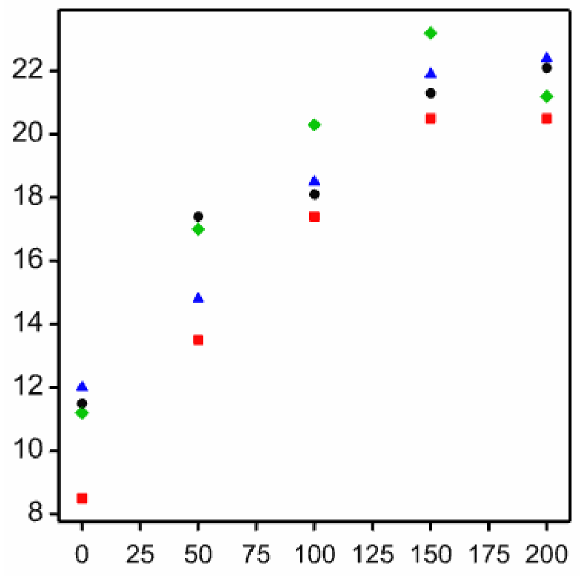

Figure S17.6.1 Crop yields vs fertilizer application coloured by block.

This extended model can then be written as

$$y_{ij} = \alpha_i + \beta \exp(-\gamma x_{ij}) + e_{ij} .$$

Fitting this model gives the summary ANOVA table in Table S17.6.1, and the model accounts for 94.4% of the variation (adjusted R^2 = 0.944). There are no obvious problems shown by the residual plots (Figure S17.6.2), but plotting the residuals against fertilizer application rates (Figure S17.6.3a) suggests some lack of fit for the 150 kg/ha rate. This is also apparent in a plot of the fitted model with the data (Figure S17.6.3b); the fitted curves show an increase from 150-200 kg/ha whereas the data shows no increase. We can formally test for lack of fit by fitting a single factor model, with a separate group for each fertilizer application; the summary ANOVA table for this model is Table S17.6.2. We deduce that PESS = 10.40, PEDF = 12, PEMS = 0.8666. We can calculate the LoFSS as the difference between the ResSS for the exponential model (= 14.98) and the PESS (= 10.40). The LoFDF is similarly the difference between the ResDF for the exponential model (= 14) and the PEDF (= 12). The F-statistic for testing LoF takes the form

$$F = \frac{(\text{ResSS} - \text{PESS}) / (\text{ResDF} - \text{PEDF})}{\text{PEMS}} = \frac{(14.98 - 10.40) / (14 - 12)}{0.8666} = 2.643$$

with 2 and 12 df. There is no evidence (P = 0.112) of lack of fit to the exponential model.

Table S17.6.1 Summary ANOVA table for exponential model for plot yields.

Term	df	SS	Mean square	Variance ratio	*P*
Model	5	347.33	69.46	64.92	< 0.001
Residual	14	14.98	1.07		
Total	19	362.31			

Table S17.6.2 Summary ANOVA table for single factor model for plot yields.

Term	df	SS	Mean square	Variance ratio	*P*
Model	7	351.91	50.272	58.01	< 0.001
Residual	12	10.40	0.867		
Total	19	362.31			

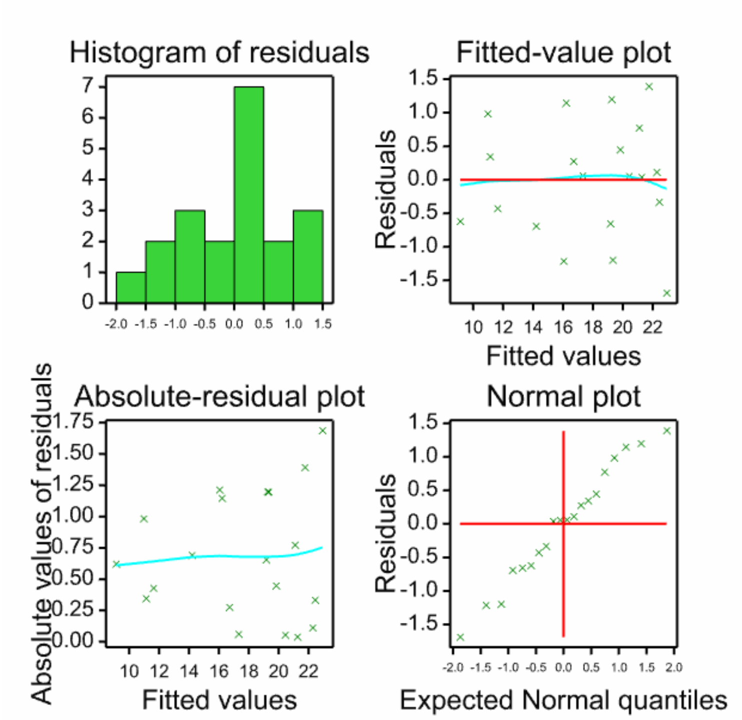

Figure S17.6.2 Composite set of residual plots based on standardized residuals from exponential model for plot yields with separate block intercepts.

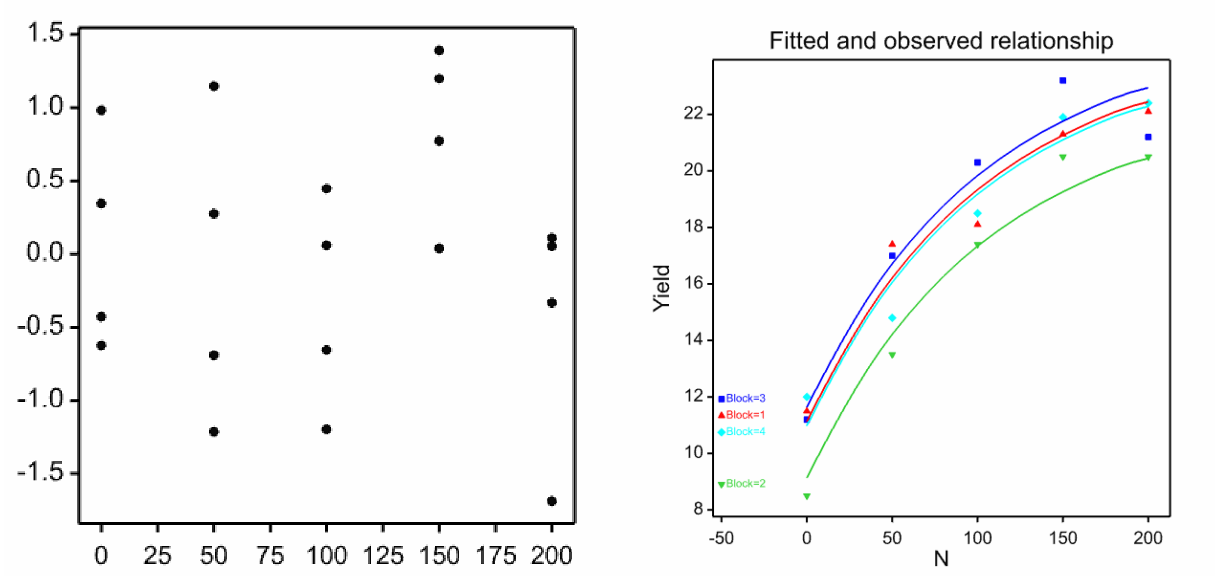

Figure S17.6.3 Exponential model for plot yields: (a) residuals plotted against fertilizer application rate; (b) fitted exponential curves with data.

We notice that the single factor model used to calculate pure error quantities uses only 2 df more than the exponential model. This happens because the exponential model has three parameters to model response across fertilizer application rates and there are only five different rates. If the intention of the experiment had been to explore the fertilizer response in detail, it would have been better to use more application rates. In order to accommodate the plateau observed at the higher application rates, we might consider adding a linear component to the exponential model, and repeating the above procedure. However, this means an additional parameter, giving only 1 fewer than the saturated single factor model, so there is a danger of over-fitting.