Solution to Exercise 15.1 (Version 1, 1/1/16)

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 15.1

The biomass data (wet weights in g) for all four sites in the study described in Exercise 12.2 (and Exercise 13.4) are held in file ALLSITES.DAT (variate *ID*, factor Site, variates *Year*, *WetWeight*). Use regression with groups on this combined data set to investigate whether the trend found at Hereford is the same as for the other three sites. Present a summary of the results from your analysis.

Solution 15.1

The combined data set is plotted against year with the log-transformation used in previous questions, i.e. log(WetWeight+0.5) in Figure S15.1.1. The relationship within each site appears approximately linear, but variation with Starcross and Wye appears smaller than that at Rothamsted and (especially) Hereford. We will perform a regression with groups to investigate differences in trend between sites, and we will investigate whether variance heterogeneity exists in the residuals from this analysis.

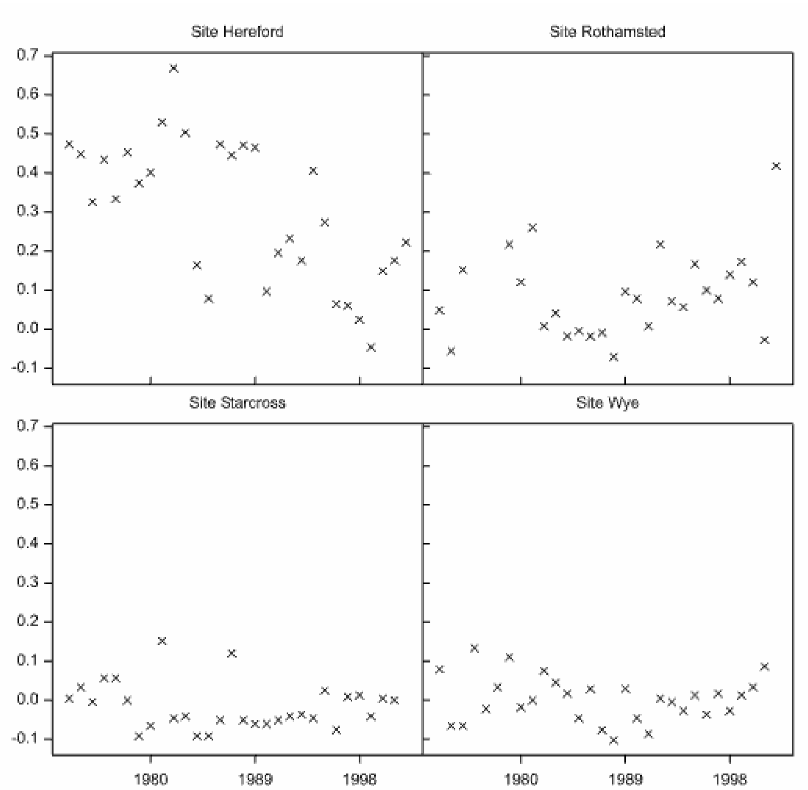

Figure S15.1.1. Log(Wet weight + 0.5) plotted against year for 4 UK sites.

We start by fitting the most complex (separate lines) model, and will then progressively simplify this if possible. This initial model is written in symbolic form as:

Response: *logWt*
Explanatory component: *[1]* + *Year**Site

A composite set of residual plots are shown in Figure S15.1.2. There is a suggestion of some variance heterogeneity but it is not marked. The incremental ANOVA table for this model is shown in Table S15.1.1, and shows that the separate lines model cannot be simplified.

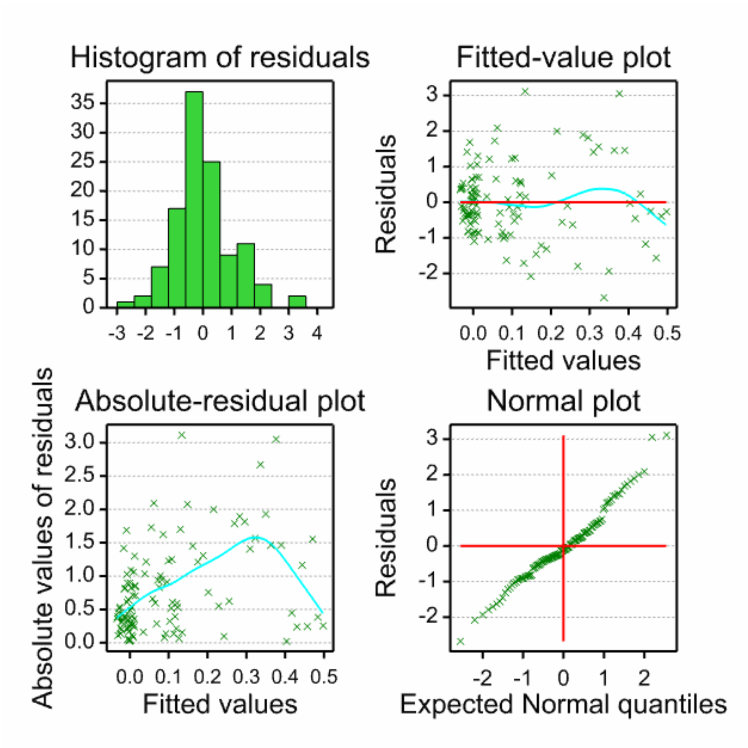

Figure S15.1.2. Composite set of residual plots (based on standardized residuals) from separate lines model for log-transformed wet weights.

Table S15.1.1 Sequential ANOVA table for separate lines models for response variate *logWt*.

Term added	Incremental df	Incremental SS	Mean square	Variance ratio	*P*
+ *Year*	1	0.0767	0.0767	7.980	0.006
+ Site	3	1.9170	0.6390	66.457	< 0.001
+ *Year*.Site	3	0.3374	0.1125	11.695	< 0.001
Residual	107	1.0288	0.0096		
Total	114	3.3599			

The separate lines model is therefore our candidate predictive model, so we investigate a little further. There is a slight suggestion of skewness in the histogram and of variance increasing with the mean in Figure S15.1.2. We can repeat the fitted values and absolute residuals plots, colouring residuals by Site (Figure S15.1.3) to see whether there is homogeneity of residuals across sites. It seems clear that variation at Starcross (green triangles) and Wye (purple crosses) is lower than variation at Rothamsted (red squares) and Hereford (blue lozenges), and this does appear to be related to predicted values. However, alternative transformations (eg. cube root) are even less successful in stabilising the variance. We might think of using a different probability distribution but there is no obvious non-Normal distribution for this type of continuous data. If we compare these results with analysis of the untransformed wet weights (not shown) then it clear that the current analysis gives a large improvement.

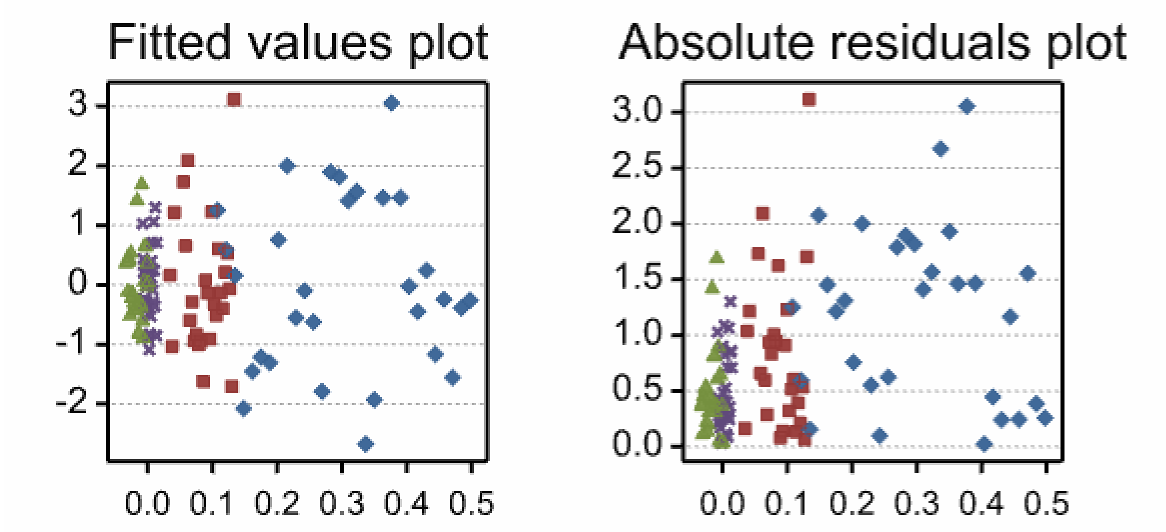

Figure S15.1.3. (a) Fitted values plot and (b) absolute residuals plot coloured by site (Hereford = blue lozenges; Rothamsted = red squares; Starcross = green triangles; Wye = purple crosses).

We proceed to summarize results from analysis of the log-transformed wet weights. The model accounts for 67.4% of the variation in log-transformed wet weights (adjusted R^2 = 0.674). The parameter estimates are shown in Table S15.1.2 (using the first-level-zero parameterization defined in Section 15.1.2.3) and these estimates can be used to construct the following predictive relationships:

Hereford: $logWt(Year) = 27.02 - 0.0134 \times Year$
Rothamsted: $logWt(Year) = -6.69 + 0.0034 \times Year$
Starcross: $logWt(Year) = 2.38 - 0.0012 \times Year$
Wye: $logWt(Year) = 1.57 - 0.0008 \times Year$

Considering these combined estimates of slope, only the slope for Hereford is significantly different from zero. We can conclude that there is a decrease in (log-transformed) wet weights over time at Hereford, but not at the other sites.

Table S15.1.2 Parameter estimates with standard errors (SE), t-statistics (t) and observed significance levels (*P*) for a separate lines model for log-transformed wet weights (*logWt*) with explanatory variate *Year* and factor Site.

Term	Parameter	Estimate	SE	t	*P*
[1]	α_1	27.021	4.1109	6.573	< 0.001
Year	β_1	-0.0134	0.00207	-6.499	< 0.001
Site 1 (H)	ν_1	0	-	-	-
Site 2 (R)	ν_2	-33.707	6.0949	-5.530	< 0.001
Site 3 (S)	ν_3	-24.644	5.9666	-4.130	< 0.001
Site 4 (W)	ν_4	-25.456	5.9666	-4.266	< 0.001
Year.Site 1	η_1	0	-	-	-
Year.Site 2	η_2	0.0168	0.00307	5.496	< 0.001
Year.Site 3	η_3	0.0122	0.00300	4.076	< 0.001
Year.Site 4	η_4	0.0127	0.00300	4.216	< 0.001