

Solution to Exercise 6.2 (Version 1, 26/09/14)

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

© S J Welham, S A Gezan, S J Clark & A Mead, 2014.

Exercise 6.2*

A pilot study investigated the pattern of an insect pest (beetle) entering a susceptible field crop. It was suspected that the beetles entered the crop from the edge of the field and then progressed towards the centre. One field was surveyed periodically and, once the beetles were present in reasonable numbers, a transect was taken from the edge towards the centre of the field with samples taken at 2 m intervals. At each distance, beetle counts were made from four randomly selected plants, giving replicate measurements at each distance. The file TRANSECT.DAT contains the unit numbers (*DPlant*), distances (factor *fDist*) and beetle counts (variate *Count*). Analyse these data, using a transformation if necessary, to investigate whether there is any evidence that beetle numbers vary among sampling distances. What other hypotheses might you like to test?

Solution 6.2

The data are plotted in Figure S6.2.1. The plot suggests variance heterogeneity: the range of observations is wider for 0 and 6 m (range = 17 and 16 beetles, respectively) than for the other distances (maximum range = 5 beetles).

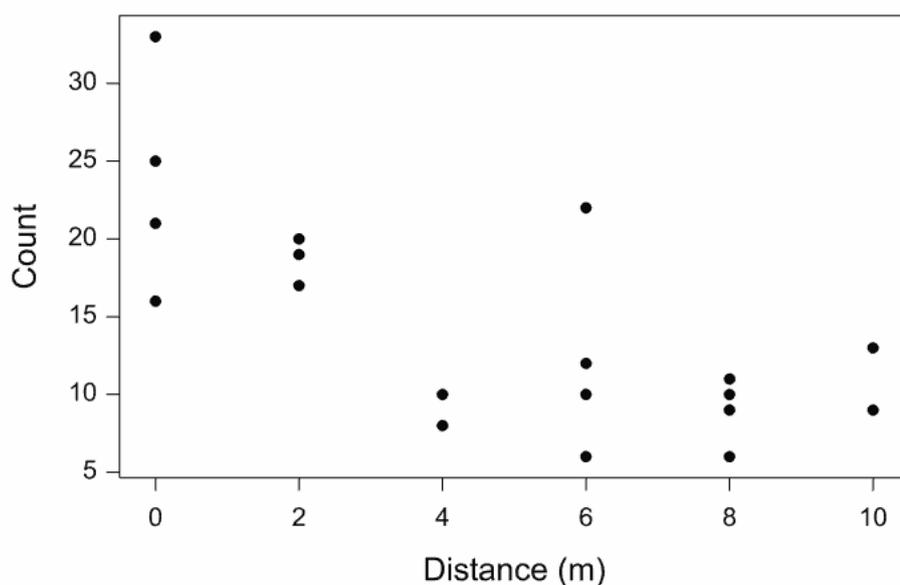


Figure S6.2.1. Scatter plot of beetle counts taken along a transect.

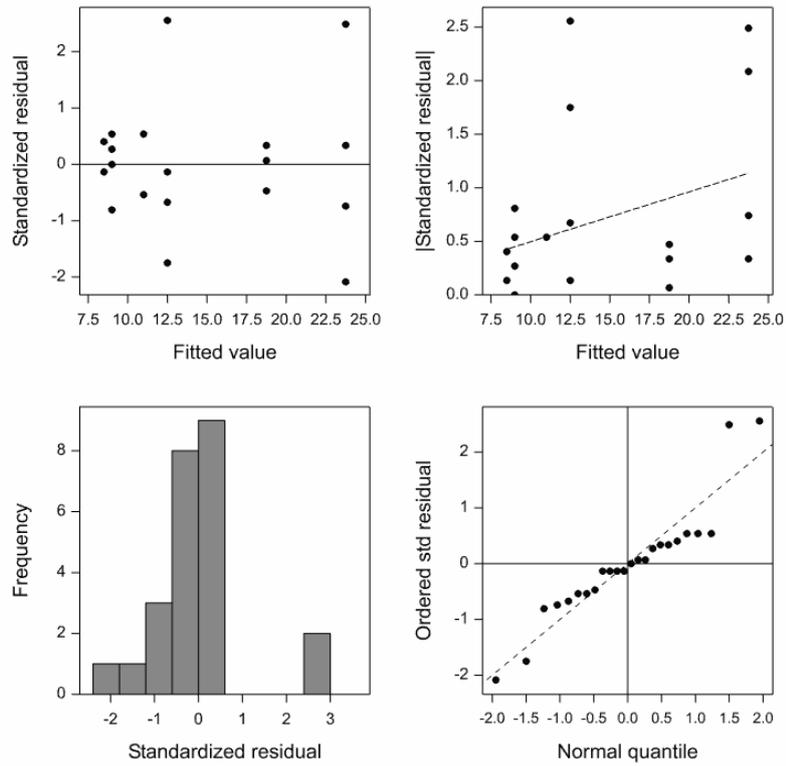


Figure S6.2.2. Composite set of residual plots based on standardized (std) residuals obtained from analysis of the observed beetle counts.

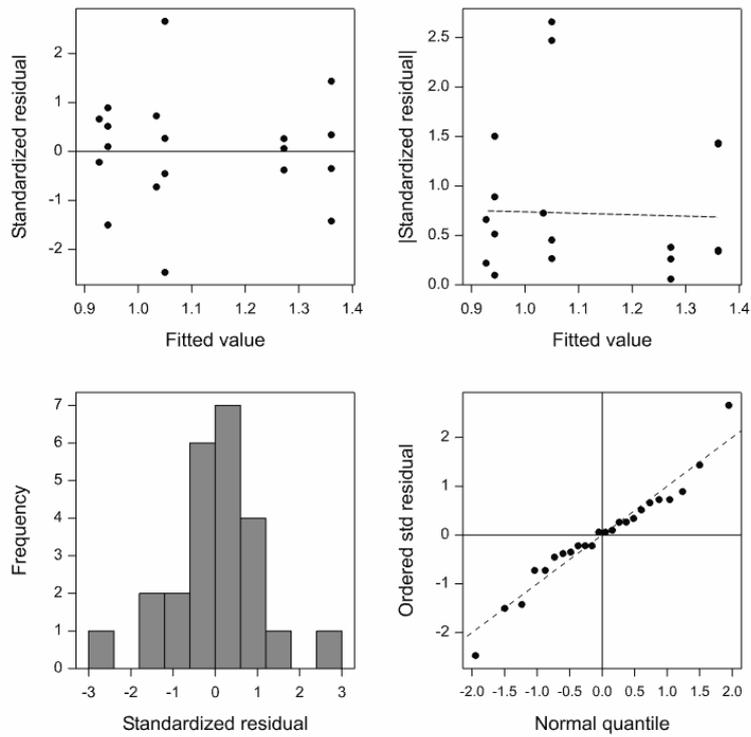


Figure S6.2.3. Composite set of residual plots based on standardized (std) residuals obtained from analysis of the logged beetle counts.

We can write a single factor model for the beetle counts as

Response variable: *Count*
 Explanatory component: [1] + fDist

A composite set of residual plots from a one-way analysis of variance of this model is shown in Figure S6.2.2. The fitted values plot (top left) and the absolute residuals plot (top right) both exhibit the same pattern of variance heterogeneity as seen in the plot of the raw data. The histogram of residuals (bottom left) is skewed and asymmetric, and the normal plot (bottom right) is non-linear at both ends; the residuals do not appear to follow a normal distribution. We seek a transformation to make the data consistent with a normal distribution with a common variance.

As these data are counts, we might expect a log transformation to stabilise the variance. Here we use common logarithms in the transformation, i.e. $\log_{10}(\text{Count})$, and use this in place of the original counts. The transformed data are plotted in Figure 12.9. The composite set of residual plots from analysis of the log-transformed data is shown in Figure S6.2.3. These are generally improved. Overall the variance appears more equal across the different distances. The histogram is more symmetric and the normal plot is straighter. We conclude that the log scale is more appropriate than the untransformed scale and interpret the analysis on this scale.

The ANOVA table for the \log_{10} -transformed counts is shown in Table S6.2.1. The variance ratio ($F_{5,18} = 7.871$) is larger than the 0.1% critical value of the F-distribution ($F_{5,18}^{0.001} = 6.808$). We therefore reject the null hypothesis (with $P < 0.001$) and conclude that there are differences among the population means for logged beetle counts taken at different distances into the field.

Table S6.2.1 ANOVA table for logged beetle counts at distances along a transect.

Source of variation	df	Sum of squares	Mean square	Variance ratio	P-value
fDist	5	0.6351	0.1270	7.871	< 0.001
Residual	18	0.2905	0.0161		
Total	23	0.9256			

The predicted means are given in Table S6.2.2. Beetles are more abundant near the edge of the field (0 and 2 m) than at distances further into the field (4–10 m), with a statistically significant decrease in population mean logged count (compared to the LSD) between 2 and 4 m. We might wish to formally investigate the hypothesis that counts decrease linearly as a function of distance into the field (with a null hypothesis of no trend). We do this using simple linear regression in Example 12.2*. Examination of the pattern of predicted means (using the LSD) suggests that there is no difference between population means near the edge of the field (0 and 2m, difference = 0.089) or between in-field logged counts (4, 6, 8 and 10m, max difference = 0.123). Hence, we might investigate a hypothesis that there are higher numbers at the edge of the field (0-2m) than at in-field positions (≥ 4 m), with no additional systematic variation. This can be done using nested factors (see Section 8.4) – we leave this as a further exercise for the reader.

* Note: We also ask you to re-analyse these data using a GLM in Exercise 18.4.

Table S6.2.2 Predicted mean logged beetle counts at distances along a transect with SED and LSD (18 df). Back-transformation of each mean given in parentheses.

Distance along transect, m							
0	2	4	6	8	10	SED	LSD
1.361	1.272	0.927	1.050	0.943	1.034	0.0898	0.1887
(22.95)	(18.72)	(8.46)	(11.22)	(8.78)	(10.82)	-	-