

Solution to Exercise 18.4 (Version 1, 22/8/15)

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

© S J Welham, S A Gezan, S J Clark & A Mead, 2015.

Exercise 18.4*

Example 12.2 analysed a set of insect counts from a transect sample and we used a log transformation to deal with variance heterogeneity. Repeat the analysis (the data are in file TRANSECT.DAT) using a suitable GLM and compare your results with the original analysis. Which analysis do you think is more appropriate?

Data 18.4 (TRANSECT.DAT)

Plant number (DPlant), distance from edge of crop (Distance, m), factor labelling distance groups (fDist), and count of pollen beetles (Count).

DPlant	Distance	fDist	Count	DPlant	Distance	fDist	Count
1	0	1	21	13	6	4	12
2	0	1	33	14	6	4	10
3	0	1	25	15	6	4	6
4	0	1	16	16	6	4	22
5	2	2	19	17	8	5	10
6	2	2	20	18	8	5	6
7	2	2	17	19	8	5	9
8	2	2	19	20	8	5	11
9	4	3	8	21	10	6	9
10	4	3	10	22	10	6	9
11	4	3	8	23	10	6	13
12	4	3	8	24	10	6	13

Solution 18.4

We might expect the beetle counts to have a Poisson distribution. For comparison with the analysis in Example 12.2, we will fit a GLM with a Poisson distribution and log link function. As in Example 12.2, we will test for lack of fit. This model can be written in symbolic form as

Response variable: *Count*
Probability distribution: Poisson
Link function: log
Explanatory component: $[1] + Distance + fDist$

The ANODEV table for this model is Table S18.4.1. We first check for over-dispersion: the residual deviance is 20.805 with 18 df, and there is no evidence of over-dispersion when compared to a chi-square distribution with 18 df ($P = 0.289$).

Table S18.4.1 Sequential ANODEV table for GLM with Poisson distribution and log link for beetle counts in terms of distance, testing for lack of fit and assuming no over-dispersion.

Source of variation	df	Deviance	Mean deviance (Chi-squared prob.)	<i>P</i>
+ <i>Distance</i>	1	32.867	32.867	< 0.001
+ fDist	4	17.063	4.266	0.002
Residual	18	20.805	1.156	
Total	23	70.735		

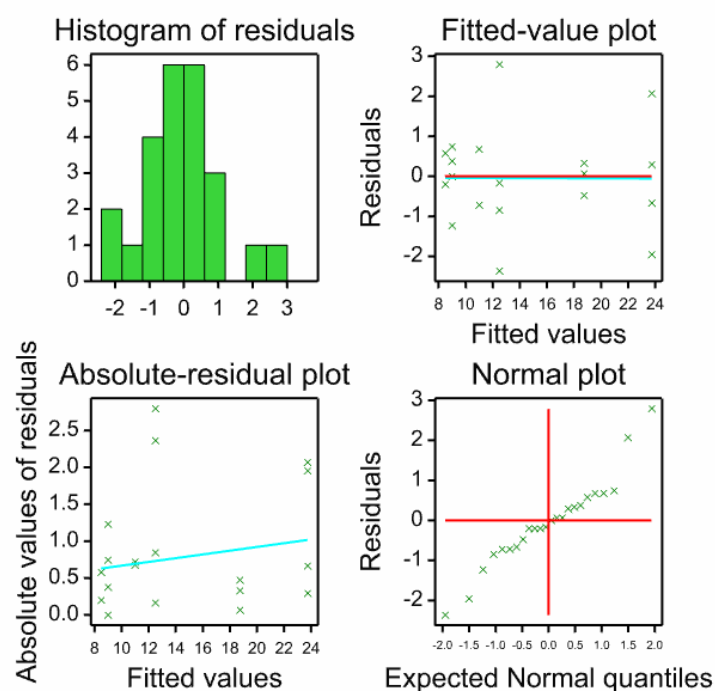


Figure S18.4.1. Composite set of residual plots for GLM with Poisson distribution and log link for beetles counts, testing for lack of fit.

A composite set of residual plots is in Figure S18.4.1. There are two large residuals which might be inspected as potential outliers, but otherwise the plots show no great cause for concern given the small number of observations. The ANODEV table shows very strong evidence of lack of fit to the linear model on the log scale. This matches the conclusion from analysis of the log-transformed data. Figure S18.4.2 shows the fitted means from these two analyses with the observed counts, and it is clear that there is little difference between them. The GLM fitted means match the arithmetic means of the treatment groups whereas the regression on log-transformed counts yields the geometric means of the treatment groups. The difference is only visible in the two groups (0m and 6m) with greater spread. In principle, we might prefer the GLM as its assumptions are more likely to match the underlying properties of the data. In practice, as we have seen, the choice makes little difference; this will not always be the case.

