Solution to Exercise 18.9 (Version 1, 30/8/15)

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

Exercise 18.9 (Data: courtesy S. Foster, Rothamsted Research)

An investigation of response to insecticide used 28 cages of clones each produced from a single aphid. There were 14 cages of each type of clone (S and R) and a target dose of active compound was applied to each cage, with the actual dose recorded. After a given period, the number of moving aphids in each cage was counted, and the clones were classified according to presence of a marker suspected to affect tolerance of the compound. File CLONE.DAT contains unit numbers (*ID*), clone type (factor Clone), marker presence (factor Marker), and the logarithm of the dose applied (variate *LogDose*) with the number of moving aphids (variate *Moving*) and total aphids (variate *Total*) in each cage. Plot the data and comment on the structure of the groups (combinations of clones and marker types). Identify and write down a parsimonious predictive model to describe the data.

Data 18.9 (CLONE.DAT)

ID	Clone	Marker	LogDose	Moving	Total	ID	Clone	Marker	LogDose	Moving	Total
1	S	-	-1.134	35	42	15	R	I	-0.343	35	94
2	S	-	-1.135	31	60	16	R	I	-0.297	11	60
3	S	-	-1.137	29	60	17	R	I	-0.478	7	54
4	S	-	-1.075	52	76	18	R	I	-0.420	23	116
5	S	-	-0.368	27	43	19	R	-	-0.014	26	80
6	S	-	-0.352	32	114	20	R	-	0.115	13	60
7	S	-	0.188	20	60	21	R	-	0.036	22	60
8	S	I	0.065	21	40	22	R	-	0.464	23	60
9	S	I	0.279	22	40	23	R	-	0.542	7	35
10	S	-	0.448	19	60	24	R	I	0.436	6	30
11	S	I	0.521	34	60	25	R	I	0.395	11	50
12	S	I	0.463	12	40	26	R	-	0.472	2	39
13	S	I	0.498	4	34	27	R	-	0.352	4	50
14	S	I	0.497	32	60	28	R	-	0.373	5	80

Solution 18.9

Figure S18.9.1 plots the proportion of moving insects against the log dose of active compound for each type of clone, with absence or presence of marker indicated. It appears that the proportion of moving insects is greater for the S-type clones, but there is no obvious relationship with log dose. We also notice that the markers are not distributed over the full range of dose within clones; in particular, there are no instances of insects with the marker at the lowest dose. There are also no resistant clones that received the lowest dose.

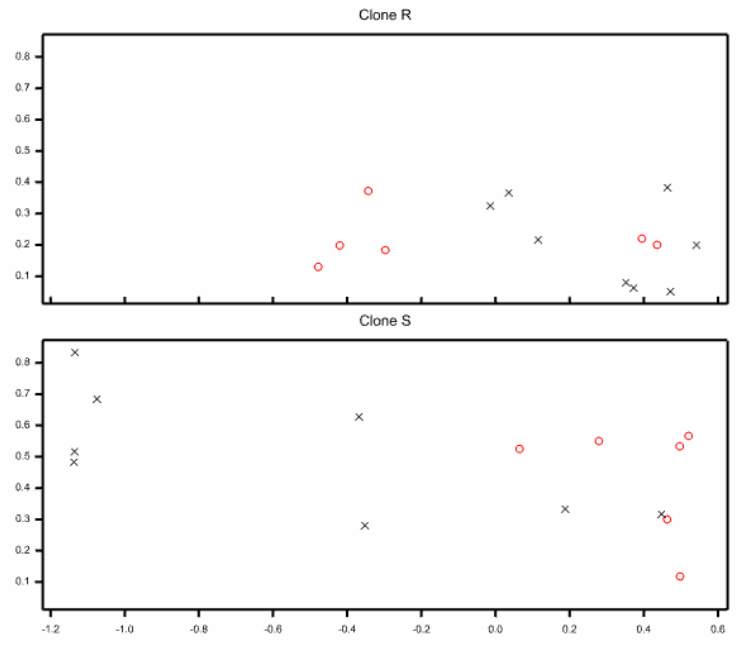

Figure S18.9.1. Proportion of moving insects plotted against log dose of compound for each type of clone (R or S), coloured by absence (×) or presence (o) of genetic marker.

The unbalanced nature of the doses across the four groups (clone type × marker) may give some ambiguity in attributing effects to treatments, and the structure is certainly not orthogonal.

We will assume a Binomial distribution for the number of moving insects out of the total in each cage at the end of the experiment. We will model the proportion of moving insects in each cage as a function of the three explanatory variables, the treatment groups (clone type and marker absence/presence) and the explanatory variate log dose of active compound. The full model can be written in symbolic form as

Response variable:	*Moving*
Probability distribution:	Binomial (number of tests = *Total*)
Link function:	logit
Explanatory component:	[1] + Clone*Marker**LogDose*

This model fits the logit-transformed mean response as a linear function of log-dose of the active compound, allowing the slope to vary between the clone × marker combinations. The summary ANODEV table from this full model is in Table S18.9.1.

Table S18.9.1 Summary ANODEV table for GLM with Binomial distribution and logit link for proportion of moving insects.

Source of variation	df	Deviance	Mean deviance	*P* (Chi-squared prob.)
Model	7	157.3	22.477	< 0.001
Residual	20	118.4	5.918	
Total	27	275.7		

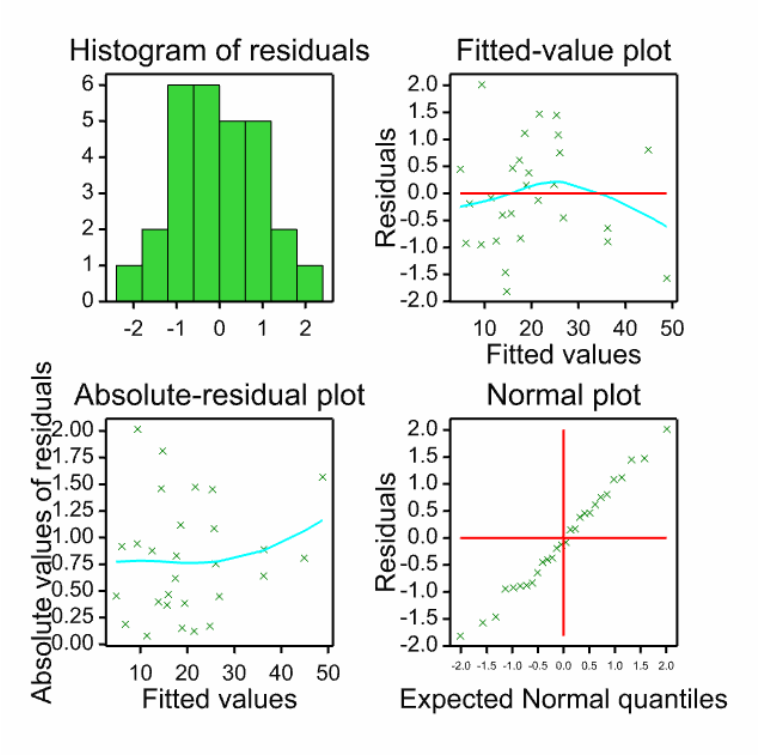

Figure S18.9.2. Composite set of residual plots based on standardized deviance residuals for GLM for proportion of moving insects with Binomial distribution and logit link.

Table S18.9.3 A sequential ANODEV table for GLM for number of emerged plants with Poisson distribution and log link.

Source of variation	df	Deviance	Mean deviance	Deviance Ratio	*P* (F prob.)
+ *LogDose*	1	51.933	51.933	8.77	0.007
+ Clone	1	82.871	82.871	14.00	0.001
+ Marker	1	1.300	1.300	0.22	0.644
+ *LogDose*.Clone	1	2.991	2.991	0.51	0.485
+ *LogDose*.Marker	1	9.978	9.978	1.69	0.209
+ Clone.Marker	1	5.008	5.008	0.85	0.369
+ *LogDose*.Clone.Marker	1	3.257	3.257	0.55	0.467
Residual	20	118.368	118.368		
Total	27	275.707			

The residual deviance (118.4) is much larger than would be expected for a chi-square distribution with 20 df ($P < 0.001$), suggesting that over-dispersion is present. We therefore re-fit the model with a dispersion parameter and examine the residuals, as in Figure S18.9.2. These plots are reasonable given the small number of observations, and so we will start to examine the model in more detail. A sequential ANODEV table is shown in Table S18.9.3, and it seems that many of the terms have small

deviance ratios. To deal with the non-orthogonal structure, we will use marginal F-tests to identify a predictive model; this process is shown in Table S18.9.4. Because the number of residual df is relatively small in the full model (ResDF = 20), we explicitly refit the model and recalculate the residual deviance at each step of the process. At the first step (Model 1), we drop the LogDose.Clone.Marker term (P = 0.467) and then refit. At the second step (Model 2), we can generate marginal F-tests for all three terms with two variables; none of these appears to explain variation in the data and we drop the term with the largest observed significance level (LogDose.Clone with P = 0.775). In model 3, we test the remaining terms with two variables and drop term LogDose.Marker (P = 0.397). In model 4, there are no interactions with *LogDose* left in the model, and so we can test this term (P = 0.015) with the Clone.Marker interaction (P = 113) and drop the latter term, leaving the three single-variable terms (Model 5). Testing these terms indicates that the Marker main effects can be omitted (P = 0.642), leaving the *LogDose* and Clone terms in the final predictive model (Model 6). This final model represents a parallel lines model, with the two clone types having separate intercepts with a common linear response (on the logit scale) to log-dose of the active compound. The parameters from the final predictive model are shown in Table S18.9.5. Note the slight discrepancy between the observed significance of the marginal F-test for LogDose in the final predictive model (P = 0.050) and that for the slope parameter (P = 0.052); this is due to the approximate nature of the parameter SEs and the marginal F-test should usually take precedence.

Table S18.9.4 Observed significance level (P) for marginal F-tests in a sequence of models for proportion of moving insects with explanatory variables Clone, Marker and *LogDose*. – = term in model but not eligible for testing, * = term omitted from model.

	P					
Term	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
LogDose	–	–	–	0.015	0.049	0.050
Clone	–	–	–	–	< 0.001	< 0.001
Marker	–	–	–	–	0.642	*
LogDose.Clone	–	0.775	*	*	*	*
LogDose.Marker	–	0.418	0.397	*	*	*
Clone.Marker	–	0.363	0.322	0.133	*	*
LogDose.Clone.Marker	0.467	*	*	*	*	*

Table S18.9.5 Parameter estimates (first-level-zero parameterization) with standard errors (SE), t-statistics (t) and observed significance level (P), for proportion of moving insects in terms of *LogDose* of active compound and Clone type (1=R or 2=S).

Term	Parameter	Estimate	SE	t	*P*
[1]	μ_1	-1.231	0.193	-6.36	< 0.001
LogDose	β	-0.479	0.234	-2.04	0.052
Clone 1	C_1	0	–	–	–
Clone 2	C_2	0.999	0.264	3.79	< 0.001

We can write the predictive model in mathematical form (with first-level-zero parameterization) as

$$\text{logit}[\hat{p}_i(l)] = \hat{\eta}_i(l) = \hat{\eta}_1 + \hat{\beta}l + \hat{C}_i,$$

where

- $p_i(l)$ is the predicted proportion of moving insects in a cage with the i^{th} clone (1=R, 2=S) and log-dose l of active compound
- $\eta_i(l)$ is the logit-transformed predicted proportion
- β is the slope of the linear response to log-dose
- C_i is the effect of the i^{th} clone type (with C_1=0)

The logit-transformed proportion of moving insects is larger for clones of type S, and decreases as the log-dose of active compound is increased for both clone types. There is no evidence that the presence or absence of the marker affects the proportion of moving insects once these factors have been taken into account. The fitted model is shown on the back-transformed scale with the observed proportions in Figure S18.9.3.

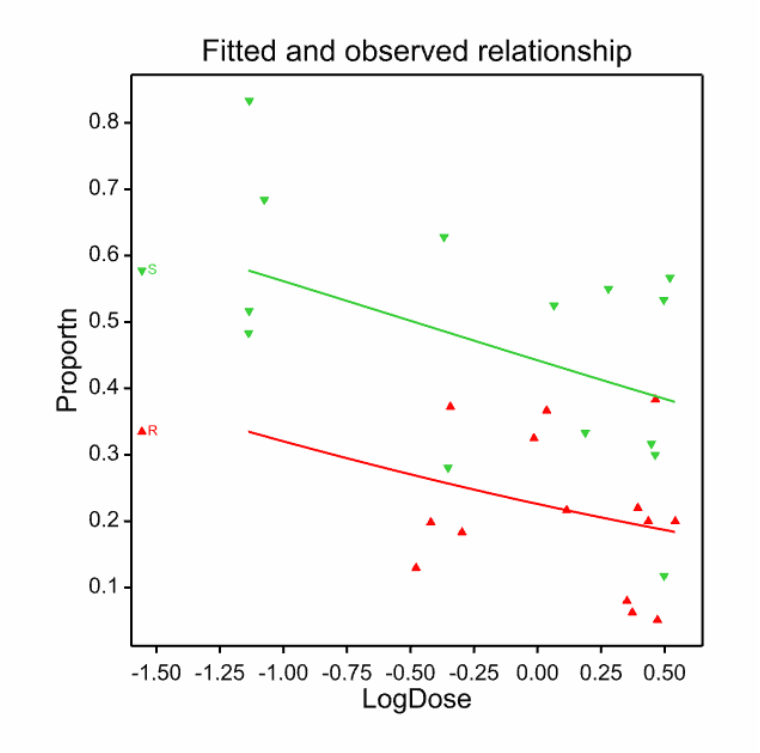

Figure S18.9.3 Fitted predictive model for proportion of moving insects in terms of clone type and log-dose of active compound.