**Solution to Exercise 18.10 (**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 18.10** (Data: courtesy Horticulture Research International)

A cage experiment was used to investigate the effect of three related insecticides on colonies of aphids with partial resistance to their common active compound. There were eight treatments: all combinations of the three insecticides or control (no insecticide) with two types of colony (susceptible or partially-resistant). The experiment was organised as a RCBD with six blocks of eight cages, and one treatment combination was allocated to each cage in each block. A colony of the designated type was reared in each cage, and the number of live aphids was counted before the insecticide treatment was applied and then two and six days after application. Both births and deaths could occur within each cage between assessments. File REPEAT.DAT holds the structural factors (*ID*, Block, Cage), treatment factors (Insecticide, Clone) and responses (variates *Pre*, *Day2*, *Day6*). First, use a GLM to analyse the numbers before the insecticide treatment is applied. Should you take account of any differences in your analysis of the post-treatment numbers? How can you do this? How does this change the interpretation of the analysis?

**Data 18.10 (REPEAT.DAT)**

| ID | Blk | Cage | I | Cl | Pre | Day2 | Day6 | ID | Blk | Cage | I | Cl | Pre | Day2 | Day6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | C | R | 60 | 111 | 220 | 25 | 4 | 1 | C | R | 75 | 134 | 238 |
| 2 | 1 | 2 | C | S | 127 | 131 | 220 | 26 | 4 | 2 | C | S | 86 | 57 | 194 |
| 3 | 1 | 3 | D | R | 64 | 30 | 27 | 27 | 4 | 3 | D | R | 69 | 32 | 12 |
| 4 | 1 | 4 | D | S | 110 | 27 | 35 | 28 | 4 | 4 | D | S | 122 | 66 | 20 |
| 5 | 1 | 5 | H | R | 118 | 75 | 121 | 29 | 4 | 5 | H | R | 185 | 88 | 251 |
| 6 | 1 | 6 | H | S | 71 | 10 | 111 | 30 | 4 | 6 | H | S | 47 | 23 | 116 |
| 7 | 1 | 7 | P | R | 66 | 69 | 62 | 31 | 4 | 7 | P | R | 104 | 26 | 40 |
| 8 | 1 | 8 | P | S | 40 | 25 | 19 | 32 | 4 | 8 | P | S | 69 | 21 | 37 |
| 9 | 2 | 1 | C | R | 54 | 152 | 156 | 33 | 5 | 1 | C | R | 89 | 421 | 444 |
| 10 | 2 | 2 | C | S | 58 | 130 | 362 | 34 | 5 | 2 | C | S | 28 | 118 | 152 |
| 11 | 2 | 3 | D | R | 76 | 60 | 110 | 35 | 5 | 3 | D | R | 207 | 19 | 83 |
| 12 | 2 | 4 | D | S | 48 | 22 | 110 | 36 | 5 | 4 | D | S | 86 | 5 | 15 |
| 13 | 2 | 5 | H | R | 130 | 113 | 101 | 37 | 5 | 5 | H | R | 67 | 35 | 25 |
| 14 | 2 | 6 | H | S | 76 | 76 | 85 | 38 | 5 | 6 | H | S | 154 | 27 | 39 |
| 15 | 2 | 7 | P | R | 93 | 77 | 185 | 39 | 5 | 7 | P | R | 101 | 36 | 46 |
| 16 | 2 | 8 | P | S | 49 | 0 | 8 | 40 | 5 | 8 | P | S | 73 | 44 | 30 |
| 17 | 3 | 1 | C | R | 94 | 175 | 292 | 41 | 6 | 1 | C | R | 78 | 222 | 121 |
| 18 | 3 | 2 | C | S | 26 | 33 | 52 | 42 | 6 | 2 | C | S | 51 | 79 | 125 |
| 19 | 3 | 3 | D | R | 121 | 73 | 60 | 43 | 6 | 3 | D | R | 54 | 37 | 18 |
| 20 | 3 | 4 | D | S | 78 | 23 | 1 | 44 | 6 | 4 | D | S | 86 | 4 | 1 |
| 21 | 3 | 5 | H | R | 73 | 74 | 56 | 45 | 6 | 5 | H | R | 60 | 30 | 105 |
| 22 | 3 | 6 | H | S | 54 | 27 | 49 | 46 | 6 | 6 | H | S | 59 | 15 | 43 |
| 23 | 3 | 7 | P | R | 25 | 10 | 32 | 47 | 6 | 7 | P | R | 105 | 66 | 151 |
| 24 | 3 | 8 | P | S | 36 | 22 | 1 | 48 | 6 | 8 | P | S | 49 | 17 | 11 |

**Solution 18.10**

In this experiment, there is a run-in period to establish the colonies before the insecticide treatments are applied. We would not expect any differences between cages assigned to the different insecticide treatments, but we might suspect that differences between the colony types might occur due to possible differences in fitness and/or fecundity. If this is the case, then we need to account for this in our analysis of post-treatment results, otherwise we may be misled as to the cause of any post-treatment differences. Figure S18.10.1 shows the pre-treatment counts within blocks. As hoped, there do not seem to be any treatment differences, but there are more instances where there are larger numbers in the cages with resistant clone types than vice versa.

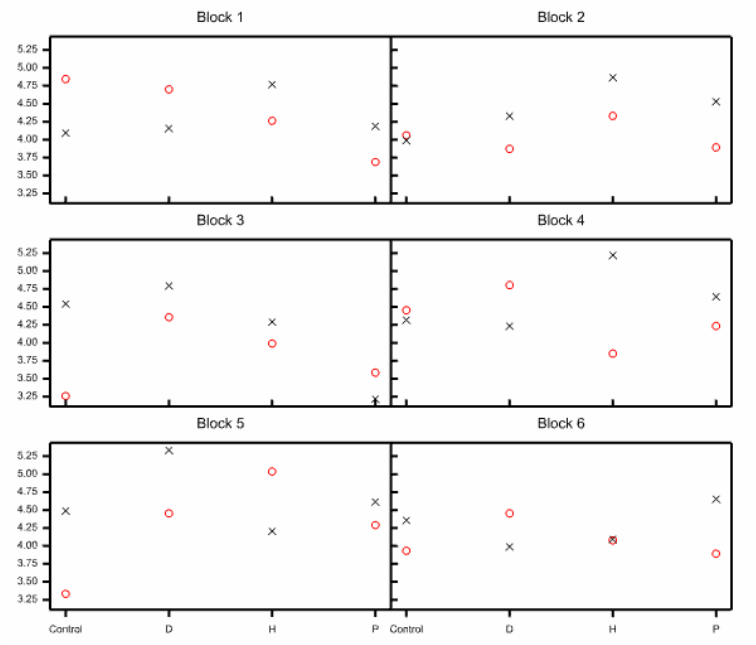


**Figure S18.10.1** Logged pre-treatment counts for each block plotted against insecticide treatment and coloured by clone type (×=R, o=S)

We therefore start by analysing the pre-treatment counts, using a Poisson distribution and a log link function. The treatment factors are crossed in a balanced 2-way factorial structure within an RCBD. As we cannot include a structural component in a GLM, we use the intra-block model by adding the blocks (as factor Block) at the start of the explanatory component. We can write this model in symbolic form as

| | |
|---|---|
| Response variable: | *Pre* |
| Probability distribution: | Poisson |
| Link function: | log |
| Explanatory component: | [1] + Block + Insecticide*Clone |

Once fit, this model shows evidence of over-dispersion (residual deviance = 496.0 with 35 df) and so we also include an estimated dispersion parameter in the model. A set of residual plots from this model are shown in Figure S18.10.2. The trend line in the absolute residual plot suggests a pattern of variance heterogeneity, but closer examination suggests this is mainly due to a few small residuals

for the lowest fitted values and a few large residuals for the highest fitted values. As there are not many observations at these extremes and the other plots seem well-behaved, we will accept these plots and proceed (with slight reservations).

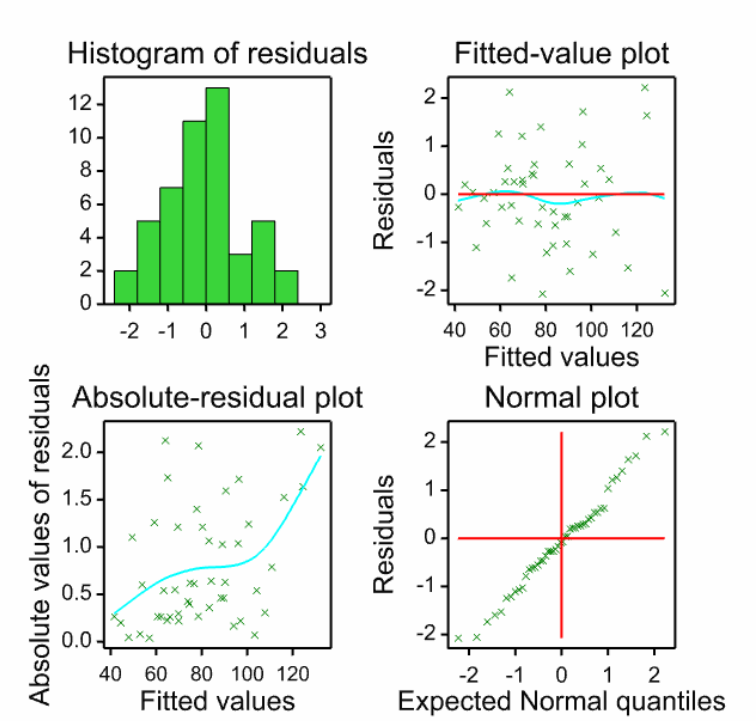


**Figure S18.10.2** Composite set of residual plots based on standardized deviance residuals for GLM for count of pre-treatment colonies with Poisson distribution and log link.

**Table S18.10.1** A sequential ANODEV table for GLM for pre-treatment colony counts with Poisson distribution and log link.

| Source of variation | df | Deviance | Mean deviance | Deviance Ratio | *P* (F prob.) |
|---|---|---|---|---|---|
| + Block | 5 | 110.52 | 22.10 | 1.56 | 0.197 |
| + Insecticide | 3 | 87.87 | 29.29 | 2.07 | 0.122 |
| + Clone | 1 | 61.24 | 61.24 | 4.32 | 0.045 |
| + Insecticide.Clone | 3 | 15.31 | 5.10 | 0.36 | 0.782 |
| Residual | 35 | 496.00 | 14.17 | | |
| Total | 47 | 770.94 | | | |

A sequential ANODEV table for the model is in Table S18.10.1, and suggests that there is no difference between insecticide treatments (as we would hope, as they have not yet been applied) but possibly a difference between clone types. To formally test this, we use a process of backwards selection with marginal F-tests. We first drop the interaction term, then the main effect for insecticide, and find some evidence that clone type affects cage numbers ($P$ = 0.047), with larger colonies in cages of resistant types.

We want to take account of these differences when we analyse the results after the treatments

are applied. However, in Figure S18.10.1 we also notice that there is substantial variation in the sizes of colonies present, on top of the overall difference between colony types. To take account of this, we might think of analysing the change in numbers before and after the treatments are applied, but we want to also take account of the heterogeneity associated with counts by using the Poisson distribution (which requires non-negative counts) and some of the changes are negative. As the Poisson distribution with log link function works in terms of multiplicative effects, it is more natural to think of analysing the ratio of counts after treatments are applied to the counts before, and this can be achieved using an offset. An offset is a term in the model without an associated parameter, and is often used in Poisson GLMs to account for differing baseline counts. We will explain further in the context of a model for the *Day2* counts. We write this model in symbolic form as

| | |
|---|---|
| Response variable: | *Day2* |
| Probability distribution: | Poisson |
| Link function: | log |
| Offset: | log(*Pre*) |
| Explanatory component: | [1] + Block + Insecticide*Clone |

In mathematical form, this model is written (using first-level-zero parameterization) as

$$\log(\mu_{ijk}) \;=\; \eta_{ijk} \;=\; \log(n_{ijk}) + \eta_{111} + B_i + I_j + C_k + (I.C)_{jk},$$

where

- $\mu_{ijk}$ is the mean response for a unit in the $i^{th}$ block with the $j^{th}$ insecticide (1=Control, 2=D, 3=H, 4=P) and $k^{th}$ colony type (1=Resistant, 2=Susceptible)
- $\eta_{ijk}$ is the log-transformed mean response for the $i^{th}$ block, $j^{th}$ insecticide and $k^{th}$ colony type
- $n_{ijk}$ is the pre-treatment count in the $i^{th}$ block, $j^{th}$ insecticide and $k^{th}$ colony type
- $B_i$ is the effect of the $i^{th}$ block (with $B_1$=0)
- $I_j$ is the effect of the $j^{th}$ insecticide treatment (1=Control, 2=D, 3=H, 4=P, with $I_1$=0)
- $C_k$ is the effect of the $k^{th}$ colony type (1=R, 2=S, with $C_1$=0)
- $(I.C)_{jk}$ is the interaction between the $j^{th}$ insecticide and the $k^{th}$ colony type, with $(I.C)_{jk}$ = 0 for $j$=1 or $k$=1

On the original scale, this model is written as

$$\mu_{ijk} \;=\; \exp(\eta_{ijk}) \;=\; n_{ijk} \times \exp[\eta_{111} + B_i + I_j + C_k + (I.C)_{jk}].$$

This models the treatment effects as multipliers of the pre-treatment numbers, and is equivalent to modelling the after:before ratio of counts, and so accounts for differing sizes of the pre-treatment colonies.

As there was over-dispersion in the pre-treatment counts, we expect it in the post-treatment counts, and so include a dispersion parameter in the model; this is confirmed by a residual deviance of 658.59 with 35 df. A set of residual plots from this model is in Figure S18.10.3. Again, at a glance the trend line in the absolute residuals plot appears to indicate variance heterogeneity, but closer inspection suggests this not reflected in the general pattern and is mainly due to two large residuals for the largest fitted values. We will accept these plots, but with perhaps a little suspicion.

A sequential ANODEV table for this model is Table S18.10.2. There is no evidence of an interaction between insecticide treatment and colony type ($P$ = 0.345) and so we drop this term from the model and re-fit. There is strong evidence that both the insecticide and colony type main effect terms influence the 2-day post-treatment colony counts, and so our final predictive model contains both of these terms plus the structural block term.

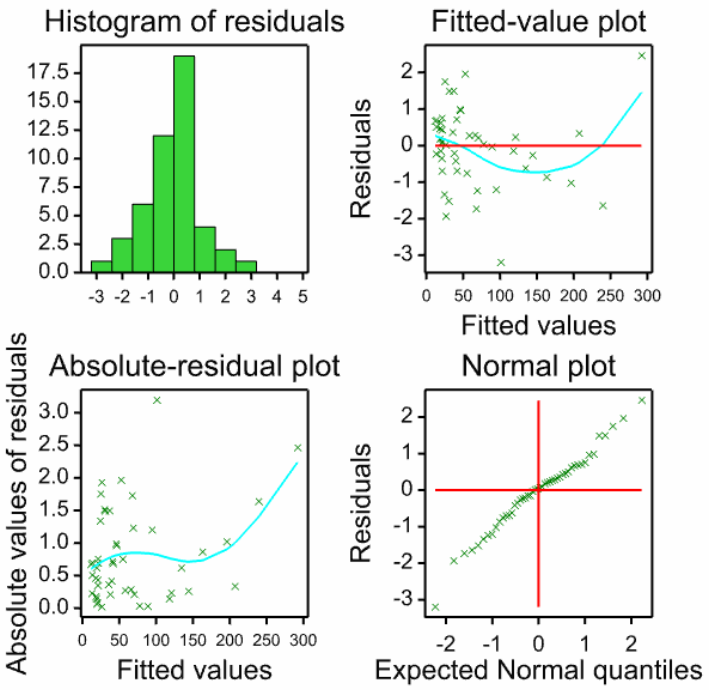


**Figure S18.10.3** Composite set of residual plots based on standardized deviance residuals for GLM for count of colonies 2 days post-treatment with pre-treatment counts as an offset, a Poisson distribution and log link.

**Table S18.10.2** A sequential ANODEV table for GLM for 2-day post-treatment colony counts with Poisson distribution and log link, using pre-treatment counts as an offset.

| Source of variation | df | Deviance | Mean deviance | Deviance Ratio | *P* (F prob.) |
|---|---|---|---|---|---|
| + Block | 5 | 109.68 | 21.94 | 1.17 | 0.345 |
| + Insecticide | 3 | 1872.43 | 624.14 | 33.17 | < 0.001 |
| + Clone | 1 | 181.66 | 181.66 | 9.65 | 0.004 |
| + Insecticide.Clone | 3 | 8.17 | 2.72 | 0.14 | 0.932 |
| Residual | 35 | 658.59 | 18.82 | | |
| Total | 47 | 2830.54 | | | |

The predictive model can be written as

$$\hat{\mu}_{ijk} = \exp(\hat{\eta}_{ijk}) = n_{ijk} \times \exp[\hat{\eta}_{111} + \hat{B}_i + \hat{I}_j + \hat{C}_k],$$

using the estimated parameters in Table S18.10.3. Predictions for a typical block can be obtained by averaging on the log-scale before back-transformation. There appear to be smaller counts in colonies of susceptible clones, and smaller counts where any of the insecticides has been applied.

**Table S18.10.3** Parameter estimates (first-level-zero parameterization) with standard errors (SE), t-statistics (t) and observed significance level (*P*), for 2-day post-treatment colony counts with pre-treatment counts as an offset and explanatory factors Block (6 levels), Insecticide (1=Control, 2=D, 3=H, 4=P) and Clone (1=R, 2=S)

| Term | Parameter | Estimate | SE | t | *P* |
|---|---|---|---|---|---|
| *[1]* | $\mu_{111}$ | 0.791 | 0.214 | 3.69 | < 0.001 |
| Block 1 | $B_1$ | 0 | – | – | – |
| Block 2 | $B_2$ | 0.484 | 0.256 | 1.89 | 0.067 |
| Block 3 | $B_3$ | 0.136 | 0.281 | 0.48 | 0.631 |
| Block 4 | $B_4$ | -0.145 | 0.277 | -0.52 | 0.603 |
| Block 5 | $B_5$ | 0.378 | 0.252 | 1.50 | 0.141 |
| Block 6 | $B_6$ | 0.174 | 0.274 | 0.64 | 0.529 |
| Insecticide 1 | $I_1$ | 0 | – | – | – |
| Insecticide 2 | $I_2$ | -1.825 | 0.234 | -7.80 | < 0.001 |
| Insecticide 3 | $I_3$ | -1.406 | 0.200 | -7.03 | < 0.001 |
| Insecticide 4 | $I_4$ | -1.500 | 0.231 | -6.50 | < 0.001 |
| Clone 1 | $C_1$ | 0 | – | – | – |
| Clone 2 | $C_2$ | -0.510 | 0.162 | -3.14 | 0.003 |

To answer the question whether there is any difference between the three insecticides, we can use the methods of Section 8.5 and construct a new factor called Control, which has value 1 for the control treatment and value 2 for the three insecticides (D, H and P). The revised model takes the symbolic form

| | |
|---|---|
| Response variable: | *Day2* |
| Probability distribution: | Poisson |
| Link function: | log |
| Offset: | log(*Pre*) |
| Explanatory component: | [1] + Block + Clone + Control/Insecticide |

and the sequential ANODEV table from this model is Table S18.10.4. The term Control estimates the difference between the control and an average over the insecticide treatments, and Control.Insecticide evaluates differences among the insecticide treatments. As we would expect from the previous analysis, there is strong evidence ($P < 0.001$) of a difference between the control and the average of the insecticide treatments. There is no evidence of any differences among the insecticide treatments ($P = 0.294$). This means that we can simplify our final model further, using a common estimate for the effect of the insecticide treatments. The parameter estimates for this simpler model are in Table S18.10.5. The estimate for insecticide treatment is close to the mean of the estimates for the three insecticide treatments separately, and the other estimates are all close to their previous values. We predict that the after:before ratio is changed (multiplied) by a factor of 0.208 when insecticide treatment is used, and by a factor of 0.597 if the clones are susceptible rather than resistant, and that these two factors operate independently. There is no evidence of any difference in response between the three insecticides.

**Table S18.10.4** A sequential ANODEV table for GLM for 2-day post-treatment colony counts with Poisson distribution and log link, using pre-treatment counts as an offset.

| Source of variation | df | Deviance | Mean deviance | Deviance Ratio | *P* (F prob.) |
|---|---|---|---|---|---|
| + Block | 5 | 109.68 | 21.94 | 1.25 | 0.305 |
| + Clone | 3 | 183.66 | 183.66 | 10.47 | 0.003 |
| + Control | 1 | 1826.05 | 1826.05 | 104.07 | > 0.001 |
| + Control.Insecticide | 3 | 44.38 | 22.19 | 1.26 | 0.294 |
| Residual | 35 | 666.76 | 17.55 | | |
| Total | 47 | 2830.54 | | | |

**Table S18.10.5** Parameter estimates (first-level-zero parameterization) with standard errors (SE), t-statistics (t) and observed significance level (*P*), for 2-day post-treatment colony counts with pre-treatment counts as an offset and explanatory factors Block (6 levels), Control (1=Control, 2=insecticide treatment) and Clone (1=R, 2=S)

| Term | Parameter | Estimate | SE | t | *P* |
|---|---|---|---|---|---|
| *[1]* | $\mu_{111}$ | 0.799 | 0.216 | 3.70 | < 0.001 |
| Block 1 | $B_1$ | 0 | – | – | – |
| Block 2 | $B_2$ | 0.497 | 0.258 | 1.93 | 0.061 |
| Block 3 | $B_3$ | 0.110 | 0.283 | 0.39 | 0.699 |
| Block 4 | $B_4$ | -0.137 | 0.279 | -0.49 | 0.626 |
| Block 5 | $B_5$ | 0.353 | 0.254 | 1.39 | 0.172 |
| Block 6 | $B_6$ | 0.171 | 0.275 | 0.62 | 0.539 |
| Control 1 | $T_1$ | 0 | – | – | – |
| Control 2 | $T_2$ | -1.568 | 0.152 | -10.32 | < 0.001 |
| Clone 1 | $C_1$ | 0 | – | – | – |
| Clone 2 | $C_2$ | -0.516 | 0.164 | -3.15 | 0.003 |

A similar procedure can be used to model the colony counts at 6 days after treatments were applied. In this case, we can again use an offset to account for pre-treatment differences between cages. If we use the pre-treatment counts as the offset, then again we are using a multiplicative model for the ratio of counts at day 6 post treatment to pre-treatment counts. Alternatively, if we are specifically interested in the day 2 to day 6 period, we might use the day 2 counts as the offset, and then we are investigating the *Day6*:*Day2* ratio.