

Solution to Exercise 18.8 (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 18.8 (Data: courtesy J. Clarkson, University of Warwick)

Data from an agronomic trial is available to assess the effect of fungicide and a biological control agent on the incidence of white rot on onions. The trial was designed as a RCBD with five blocks of 12 plots. The 12 treatments were all combinations of three varieties with presence or absence of the fungicide (two levels) and the biological control agent (BCA, two levels). File `BCA.DAT` holds the unit numbers (*ID*), structural factors (*Rep*, *Plot*), treatment factors (*Variety*, *Fungicide*, *BCA*) and the total number of plants per plot (variate *Emerged*) and number with symptoms of white rot (variate *Disease*). Use a suitable GLM to identify a predictive model for both emergence and disease incidence. Note that the number of emerged plants is a small proportion of the seeds sown (which was not counted but was constant across plots) so is small compared to the unknown upper limit. What treatment would you recommend to maximise the number of unaffected plants for each variety?

Data 18.8 (BCA.DAT)

Number of emerged (E) and diseased (D) plants per plot from a trial with 5 replicates (*Rep*) of 12 plots (*Plot*) with 3 varieties (*Var*), with or without fungicide (*Fung*) and application or not of biological control agent (*BCA*).

Rep	Plot	Var	Fung	BCA	E	D	Rep	Plot	Var	Fung	BCA	E	D
1	1	HY	No	No	103	44	3	7	RB	Yes	Yes	52	17
1	2	HY	No	Yes	49	20	3	8	RB	Yes	No	75	6
1	3	HY	Yes	Yes	48	10	3	9	Ren	No	No	91	39
1	4	HY	Yes	No	45	10	3	10	Ren	No	Yes	71	17
1	5	RB	No	No	68	29	3	11	Ren	Yes	Yes	67	12
1	6	RB	No	Yes	53	28	3	12	Ren	Yes	No	79	12
1	7	RB	Yes	Yes	38	7	4	1	HY	No	No	64	20
1	8	RB	Yes	No	48	10	4	2	HY	No	Yes	47	19
1	9	Ren	No	No	80	23	4	3	HY	Yes	Yes	40	3
1	10	Ren	No	Yes	54	25	4	4	HY	Yes	No	104	3
1	11	Ren	Yes	Yes	33	11	4	5	RB	No	No	88	52
1	12	Ren	Yes	No	66	7	4	6	RB	No	Yes	57	25
2	1	HY	No	No	59	44	4	7	RB	Yes	Yes	48	9
2	2	HY	No	Yes	48	14	4	8	RB	Yes	No	74	1
2	3	HY	Yes	Yes	47	7	4	9	Ren	No	No	84	43
2	4	HY	Yes	No	72	8	4	10	Ren	No	Yes	69	18
2	5	RB	No	No	68	49	4	11	Ren	Yes	Yes	75	1
2	6	RB	No	Yes	52	27	4	12	Ren	Yes	No	86	3
2	7	RB	Yes	Yes	37	10	5	1	HY	No	No	87	35
2	8	RB	Yes	No	47	10	5	2	HY	No	Yes	61	10

Rep	Plot	Var	Fung	BCA	E	D	Rep	Plot	Var	Fung	BCA	E	D
2	9	Ren	No	No	93	46	5	3	HY	Yes	Yes	43	3
2	10	Ren	No	Yes	66	19	5	4	HY	Yes	No	103	3
2	11	Ren	Yes	Yes	43	12	5	5	RB	No	No	82	42
2	12	Ren	Yes	No	74	14	5	6	RB	No	Yes	73	11
3	1	HY	No	No	96	54	5	7	RB	Yes	Yes	33	19
3	2	HY	No	Yes	38	15	5	8	RB	Yes	No	87	7
3	3	HY	Yes	Yes	70	3	5	9	Ren	No	No	91	10
3	4	HY	Yes	No	56	9	5	10	Ren	No	Yes	78	19
3	5	RB	No	No	87	43	5	11	Ren	Yes	Yes	67	3
3	6	RB	No	Yes	81	13	5	12	Ren	Yes	No	95	8

Solution 18.8

Analysis of Emergence

We start by considering emergence. We would usually think of emergence as having a Binomial distribution, with the number of binomial tests being equal to the number of seeds sown. However, here we cannot use this model as we do not know the number of seeds sown, only that it was much larger than the number of plants that emerged. We will therefore regard the number emerged as counts without an effective upper limit, and use a Poisson distribution. (Statistical theory shows that the Poisson distribution gives a reasonable approximation to the Binomial distribution when the binomial probability of success (p) is small and number of tests (m) is large). We need to know that the number of seeds sown was similar across plots for this to be a sensible strategy, and we are assured this was the case here. We start by plotting the data in Figure S18.8.1.

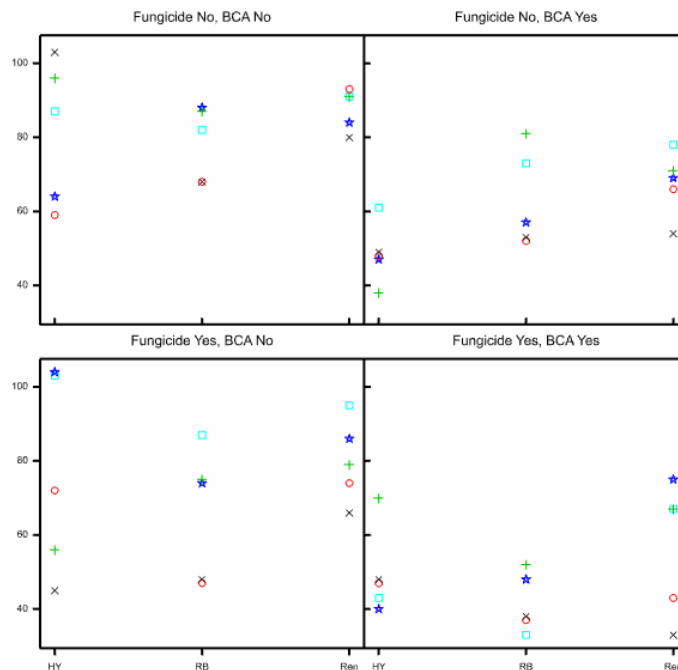


Figure S18.8.1. Observed plant counts for combinations of variety, fungicide and BCA treatments, coloured by replicate.

Figure S18.8.1 suggests that application of BCA reduces emergence, but no other effects are clear. The treatment factors are crossed in a balanced 3-way factorial structure and the design is a RCBD. As we cannot include a structural component in a GLM, we use the intra-block model by adding the blocks (here factor Rep) at the start of the explanatory component. We can write this model in symbolic form as

Response variable: *Emerged*
 Probability distribution: Poisson
 Link function: log
 Explanatory component: [1] + Rep + Variety*Fungicide*BCA

We start by fitting this model and checking for over-dispersion. The summary ANODEV table is Table S18.8.1. The residual deviance (104.3) is much larger than would be expected for a chi-square distribution with 44 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 S18.8.2.

Table S18.8.1 Summary ANODEV table for GLM with Poisson distribution and log link for number of emerged plants.

Source of variation	df	Deviance	Mean deviance	<i>P</i> (Chi-squared prob.)
Model	15	236.5	15.765	< 0.001
Residual	44	104.3	2.370	
Total	59	340.7		

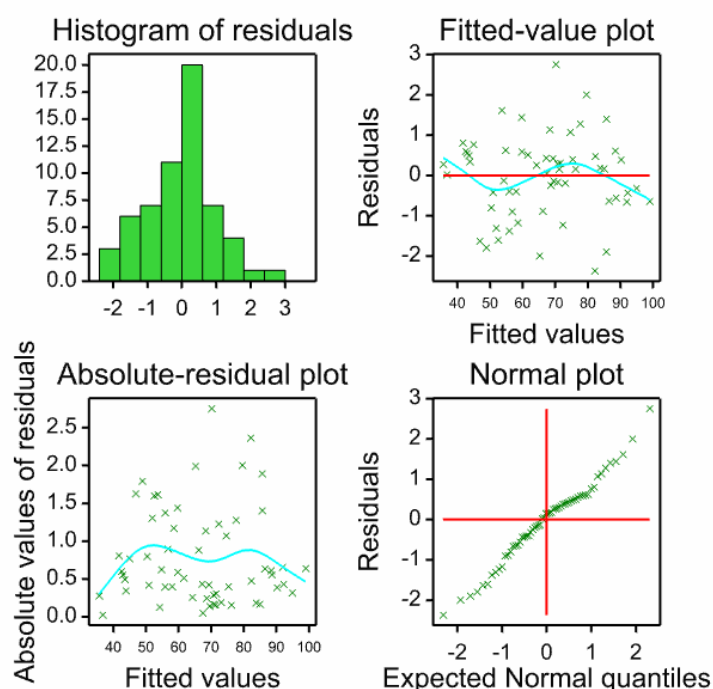


Figure S18.8.2. Composite set of residual plots based on standardized deviance residuals for GLM for number of emerged plants with Poisson distribution and log link.

The residual plots show no strong evidence that the model assumptions are not met, and so we can start to examine the fit. Table S18.8.2 shows a sequential ANODEV table for the full model. Although the design is balanced, because this is a GLM the order in which terms are fitted may affect the deviance ratios, and so we will use marginal F tests to identify the predictive model. This process is shown in Table S18.8.3. Note that here we do not recalculate the residual mean deviance at each stage of the process. Table S18.8.2 shows that we can drop the 3-way interaction ($P = 0.408$), so we drop this term and re-fit the model (Model 2), and examine the marginal F-tests for the 2-way interactions. We drop the Fungicide.BCA interaction ($P = 0.451$) and then re-examine the other two 2-way interactions (Model 3). We can then drop the Variety.BCA interaction, and re-fit (Model 4); at this stage we can examine the Variety.Fungicide interaction and the BCA main effect. We then drop the Variety.Fungicide interaction (Model 5) and test all three main effects; these all have observed significance < 0.05 and so we cannot simplify further and we have our predictive model. The parameters from this model are in Table S18.8.4.

Table S18.8.2 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.)
+ Rep	4	47.307	11.827	4.99	0.002
+ Variety	2	19.754	9.877	4.17	0.022
+ Fungicide	1	20.518	20.518	8.66	0.005
+ BCA	1	128.459	128.459	54.21	< 0.001
+ Variety.Fungicide	2	8.955	4.477	1.89	0.163
+ Variety.BCA	2	5.779	2.890	1.22	0.305
+ Fungicide.BCA	1	1.371	1.371	0.58	0.451
+ Variety.Fungicide.BCA	2	4.338	2.169	0.92	0.408
Residual	44	104.268	2.370		
Total	59	340.748			

Table S18.8.3 Observed significance level (P) for marginal F-tests in a sequence of models for plant emergence with explanatory factors Run, Variety, Fungicide and BCA. – = term in model but not eligible for testing, * = term omitted from model.

Term	P				
	Model 1	Model 2	Model 3	Model 4	Model 5
Run	–	–	–	–	–
Variety	–	–	–	–	0.022
Fungicide	–	–	–	–	0.005
BCA	–	–	–	< 0.001	< 0.001
Variety.Fungicide	–	0.169	0.163	0.163	*
Variety.BCA	–	0.316	0.305	*	*
Fungicide.BCA	–	0.451	*	*	*
Variety.Fungicide.BCA	0.408	*	*	*	*

Table S18.8.4 Parameter estimates (first-level-zero parameterization) with standard errors (SE), t-statistics (t) and observed significance level (*P*), for plant emergence with explanatory factors Rep (5 levels), Variety (1=HY, 2=RB, 3=Ren), Fungicide (1=No, 2=Yes) and BCA (1=No, 2=Yes)

Term	Parameter	Estimate	SE	t	<i>P</i>
[1]	μ_{1111}	4.2400	0.0763	55.56	< 0.001
Rep 1	R_1	0	–	–	–
Rep 2	R_2	0.0302	0.0839	0.36	0.720
Rep 3	R_3	0.2310	0.0800	2.89	0.006
Rep 4	R_4	0.1992	0.0806	2.47	0.017
Rep 5	R_5	0.2730	0.0793	3.44	0.001
Variety 1	V_1	0	–	–	–
Variety 2	V_2	-0.0253	0.0622	-0.41	0.686
Variety 3	V_3	0.1329	0.0599	2.22	0.031
Fungicide 1	F_1	0	–	–	–
Fungicide 2	F_2	-0.1436	0.0496	-2.89	0.006
BCA 1	B_1	0	–	–	–
BCA 2	B_2	-0.3618	0.0503	-7.19	< 0.001

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

$$\log(\hat{\mu}_{ijkl}) = \hat{\eta}_{ijkl} = \hat{\eta}_{1111} + \hat{R}_i + \hat{V}_j + \hat{F}_k + \hat{B}_l,$$

where

- μ_{ijkl} is the predicted mean response for a unit in the i^{th} replicate with the j^{th} variety, k^{th} level of fungicide treatment (1=no, 2=yes) and l^{th} level of BCA treatment (1=no, 2=yes)
- η_{ijkl} is the log-transformed expected mean response
- R_i is the effect of the i^{th} replicate (with $R_1=0$)
- V_j is the effect of the j^{th} variety (with $V_1=0$)
- F_k is the effect of the k^{th} level of fungicide treatment (1=no, 2=yes, with $F_1=0$)
- B_l is the effect of the l^{th} level of BCA treatment (1=no, 2=yes, with $B_1=0$)

Using this model to help us interpret the estimated parameters tells us that there is no real difference between varieties 1 and 2 (HY and RB) but variety 3 (Ren) has an increase in emergence of 0.133 units on the log scale (compared to HY), use of fungicide decreases emergence by 0.144 units on the log scale, and use of BCA decreases emergence by 0.362 units on the log scale. If we rewrite the model in terms of predicted values as

$$\hat{\mu}_{ijkl} = \exp(\hat{\eta}_{ijkl}) = \exp(\hat{\eta}_{1111} + \hat{R}_i + \hat{V}_j + \hat{F}_k + \hat{B}_l) = \exp(\hat{\eta}_{1111}) \times \exp(\hat{R}_i) \times \exp(\hat{V}_j) \times \exp(\hat{F}_k) \times \exp(\hat{B}_l),$$

we can see that use of variety 3 (Ren) increases predicted mean emergence by a multiplicative factor of $1.14 = \exp(0.133)$ when compared to HY, use of fungicide reduces predicted mean emergence by a multiplicative factor of $0.866 = \exp(-0.144)$, and use of BCA reduces predicted mean emergence by a multiplicative factor of $0.696 = \exp(-0.362)$.

To predict the mean response for a combination of treatments in an average rep, we average over replicates on the log scale and then back-transform, as in Example 18.4.

Analysis of disease incidence

We will analyse disease incidence as a proportion of the emerged plants, and use a Binomial distribution with number of diseased plants as the response and number of emerged plants as the binominal totals. Without any information to suggest otherwise, we will use a logit link function. The full model can be written in symbolic form as

Response variable:	<i>Disease</i>
Probability distribution:	Binomial (number of tests = <i>Emerged</i>)
Link function:	logit
Explanatory component:	[1] + Rep + Variety*Fungicide*BCA

The analysis otherwise follows the path used for emergence counts, so we just give the results here and do not describe the process in detail. We again find evidence of over-dispersion and so fit a dispersion parameter, and find no issues with residual plots from the full model. Selection of terms for the predictive model follows the path shown in Table S18.8.5.

Table S18.8.5 Observed significance level (*P*) for marginal F-tests in a sequence of models for disease incidence with explanatory factors Run, Variety, Fungicide and BCA. – = term in model but not eligible for testing, * = term omitted from model.

Term	<i>P</i>			
	Model 1	Model 2	Model 3	Model 4
Run	–	–	–	–
Variety	–	–	–	0.034
Fungicide	–	–	–	–
BCA	–	–	–	–
Fungicide.BCA	–	> 0.001	0.001	0.001
Variety.Fungicide	–	0.337	0.285	*
Variety.BCA	–	0.801	*	*
Variety.Fungicide.BCA	0.165	*	*	*

The final predictive model contains all three main effects, plus the Fungicide.BCA interaction, which we might write in mathematical form as

$$\text{logit}(\hat{\mu}_{ijkl}) = \hat{\eta}_{ijkl} = \hat{\eta}_{1111} + \hat{R}_i + \hat{V}_j + \hat{F}_k + \hat{B}_l + (FB)_{kl}.$$

Here we use notation as above, but with μ_{ijkl} as the expected number of diseased plants, equal to the expected proportion (p_{ijkl}) multiplied by the number of emerged plants. The estimated parameter values are in Table S18.8.6. Here we find that disease incidence is greatest for variety 2 (RB) and least for variety 3 (Ren). We find that fungicide and BCA both reduce disease incidence, but that application of both does not reduce disease incidence any further, and may be less effective than fungicide alone. The Fungicide x BCA table of predicted means on the logit scale, averaged over varieties and replicates is in Table S18.8.7, with approximate LSDs. The difference between disease incidence with fungicide and with or without BCA is only slightly smaller than the LSD value, so

there is borderline evidence that application of BCA in addition to fungicide is not beneficial, although application of BCA alone reduces disease incidence.

Table S18.8.6 Parameter estimates (first-level-zero parameterization, excluding zero terms) with standard errors (SE), t-statistics (t) and observed significance level (*P*), for plant emergence with explanatory factors Rep (5 levels), Variety (1=HY, 2=RB, 3=Ren), Fungicide (1=No, 2=Yes) and BCA (1=No, 2=Yes)

Term	Parameter	Estimate	SE	t	<i>P</i>
[1]	μ_{1111}	-0.021	0.235	-0.09	0.930
Rep 2	R_2	0.324	0.262	1.24	0.222
Rep 3	R_3	-0.186	0.257	-0.73	0.472
Rep 4	R_4	-0.332	0.266	-1.25	0.218
Rep 5	R_5	-0.672	0.270	-2.49	0.016
Variety 2	V_2	0.352	0.204	1.73	0.091
Variety 3	V_3	-0.175	0.205	-0.85	0.397
Fungicide 2	F_2	-2.051	0.251	-8.15	<0.001
BCA 2	B_2	-0.655	0.203	-3.23	0.002
Fungicide 2.BCA 2	$(FB)_{22}$	1.263	0.366	3.45	0.001

Table S18.8.7. Predicted means on logit scale from predictive model for disease incidence with LSDs.

Index	Fungicide	BCA	Prediction	Approximate LSD for comparison		
				vs 1	vs 2	vs 3
1	No	No	-0.179	-		
2	No	Yes	-0.834	0.4069	-	
3	Yes	No	-2.229	0.5051	0.5407	-
4	Yes	Yes	-1.621	0.4969	0.5338	0.6111

Analysis of number of unaffected plants

In order to recommend a treatment to maximise the number of unaffected plants, we should also analyse this variable. Since the number of emerged plants differs between treatments, we cannot deduce this from analysis of disease incidence. So we will analyse the variable

$$\text{Healthy} = \text{Emerged} - \text{Disease}.$$

As for the analysis of emergence, we will assume a Poisson distribution and log link. Again, we follow the same path of analysis, and arrive at a predictive model containing the same terms as for disease incidence (see Table S18.8.8). The estimated parameters are in Table S18.8.9, with predictions on the log scale in Table S18.8.10. It appears that the fungicide-only treatment maximises the number of healthy plants.

Table S18.8.8 Observed significance level (*P*) for marginal F-tests in a sequence of models for number of unaffected plants with explanatory factors Run, Variety, Fungicide and BCA. – = term in model but not eligible for testing, * = term omitted from model.

Term	<i>P</i>			
	Model 1	Model 2	Model 3	Model 4
Run	–	–	–	–
Variety	–	–	–	0.010
Fungicide	–	–	–	–
BCA	–	–	–	–
Fungicide.BCA	–	0.012	0.011	0.011
Variety.Fungicide	–	0.282	0.261	*
Variety.BCA	–	0.913	*	*
Variety.Fungicide.BCA	0.157	*	*	*

Table S18.8.9 Parameter estimates (first-level-zero parameterization, excluding zero terms) with standard errors (SE), t-statistics (t) and observed significance level (*P*), for number of unaffected plants with explanatory factors Rep (5 levels), Variety (1=HY, 2=RB, 3=Ren), Fungicide (1=No, 2=Yes) and BCA (1=No, 2=Yes)

Term	Parameter	Estimate	SE	t	<i>P</i>
[1]	μ_{1111}	3.546	0.130	27.17	< 0.001
Rep 2	R_2	-0.033	0.136	-0.24	0.809
Rep 3	R_3	0.301	0.126	2.39	0.021
Rep 4	R_4	0.327	0.126	2.60	0.012
Rep 5	R_5	0.460	0.122	3.76	< 0.001
Variety 2	V_2	-0.127	0.098	-1.30	0.198
Variety 3	V_3	0.169	0.901	1.86	0.069
Fungicide 2	F_2	0.403	0.103	3.93	< 0.001
BCA 2	B_2	-0.079	0.115	-0.69	0.492
Fungicide 2.BCA 2	$(FB)_{22}$	-0.408	0.156	-2.62	0.012

Table S18.8.10. Predicted means on log scale from predictive model for number of unaffected plants with LSDs.

Index	Fungicide	BCA	Prediction	Approximate LSD for comparison		
				vs 1	vs 2	vs 3
1	No	No	3.770	-		
2	No	Yes	3.691	0.2304	-	
3	Yes	No	4.174	0.2062	0.2113	-
4	Yes	Yes	3.686	0.2307	0.2352	0.2116