Chapter 6
Generalized Linear Mixed Models for Proportions and Percentages

6.1 Response Variables as Ratios and Percentages

In this chapter, we will review generalized linear mixed models (GLMMs) whose response can be either a proportion or a percentage. For proportion and percentage data, we refer to data whose expected value is between 0 and 1 or between 0 and 100. For the remainder of this book, we will refer to this type of data only in terms of proportion, knowing that it is possible to change it to a percentage scale only when multiplying it by 100. Proportions can be classified into two types: discrete and continuous. Discrete proportions arise when the unit of observation consists of N distinct entities, of which individuals have the attribute of interest "y". N must be a nonnegative integer and "y" must be a positive integer; here, $y \leq N$. Therefore, the observed proportion must be a discrete fraction, which can take values $\frac{0}{N}, \frac{1}{N}, \cdots, \frac{N}{N}$. A binomial distribution is the sum of a series of m independent binary trials (i.e., trials with only two possible outcomes: success or failure), where all trials have the same probability of success. For binary and binomial distributions, the target of inference is the value of the parameter such that $0 \leq E\left(\frac{y}{N}\right) = \pi \leq 1$. Continuous proportions (ratios) arise when the researcher measures responses such as the fraction of the area of a leaf infested with a fungus, the proportion of damaged cloth in a square meter, the fraction of a contaminated area, and so on. As with the binomial parameter π, the continuous rates (fractions) take values between 0 and 1, but, unlike the binomial, the continuous proportions do not result from a set of Bernoulli tests. Instead, the beta distribution is most often used when the response variable is in continuous proportions. In the following sections, we will first address issues in modeling when we have binary and binomial data. When the response variable is binomial, we have the option of using a linearization method (pseudo-likelihood (PL)) or the Laplace or quadrature integral approximation (Stroup 2012).

J. Salinas Ruíz et al., *Generalized Linear Mixed Models with Applications in Agriculture and Biology*, https://doi.org/10.1007/978-3-031-32800-8_6

6.2 Analysis of Discrete Proportions: Binary and Binomial Responses

A binomial distribution is the number of successes from a series of N independent binary trials – Bernoulli trials (i.e., trials with two possible outcomes: success or failure), where all trials have the same probability of success. In the context of a GLMM, there are N binomial responses, each of which is the result of binary trials. The ith response consists of two pieces of information: the number of trials n_i and the number of successes y_i, as shown in the following example.

6.2.1 Completely Randomized Design (CRD): Methylation Experiment

An agent to induce demethylation is applied to plants; this agent converts methylated nucleotides to their unmethylated forms, thus causing epigenetic changes that produce or induce abnormal phenotypes such as deformation or stunting (Amoah et al. 2008). A pilot study was implemented to investigate the relationship between the dose of the demethylating agent and the observed proportion of plants with a normal phenotype. Seeds were treated with the demethylating agent at six different doses, including the control. Plants were sown in trays, with each tray containing seeds previously treated with the same dose of the demethylating agent. Each dose was replicated 4 times: 2 with 60 plants and 2 with 100 plants. The trays were allocated following a completely randomized design (CRD). The plants with a normal phenotype in each tray are shown (in Table 6.1) with the number of plants per tray (N). The notation 59(60) indicates that 59 normal plants were found out of 60 plants under study. In the same way, the notation 14(100) indicates that 14 normal plants were found out of 100 plants under study.

The sources of variation and degrees of freedom (DFs) for this experiment are shown in Table 6.2.

Table 6.1 Number of normal plants out of a total of N plants per tray and dose of the demethylating agent

Dose					
0	0.01	0.1	0.5	1.0	1.5
59(60)	58(60)	54(60)	4(60)	3(60)	3(60)
58(60)	59(60)	53(60)	11(60)	2(60)	3(60)
99(100)	98(100)	88(100)	14(100)	2(100)	1(100)
98(100)	99(100)	87(100)	15(100)	1(100)	3(100)

Table 6.2 Sources of variation and degrees of freedom

Sources of variation	Degrees of freedom
Dose	$t - 1 = 6 - 1 = 5$
Error	$t(r - 1) = 6 \times (4 - 1) = 18$
Total	$t \times r - 1 = 6 \times 4 - 1 = 23$

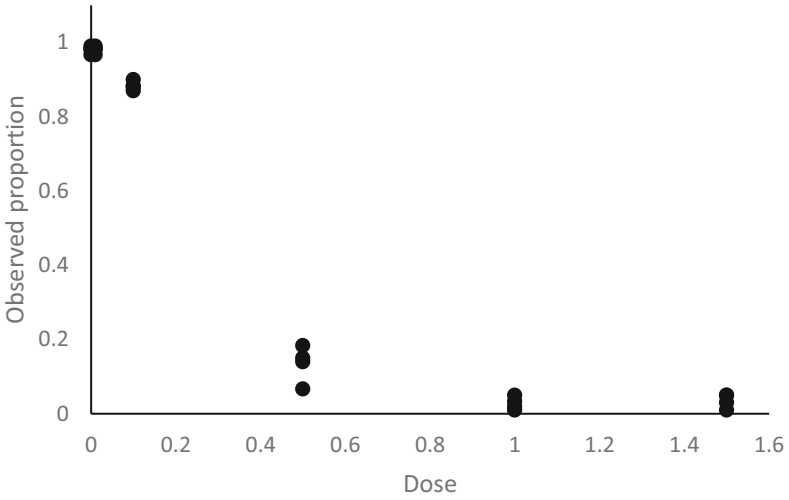

Fig. 6.1 Effect of the demethylating agent on the proportion of normal plants

The statistical model of a completely randomized design (CRD) is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where y_{ij} is the number of observed normal plants in the tray j ($j = 1, 2, 3, 4$) at the dose i ($i = 1, 2, \cdots, 6$), μ is the overall mean, τ_i is the effect of dose i of the demethylating agent, and ε_{ij} are non-normal errors.

The expected value (normal plants) of a set of tests n_i follows a binomial distribution $y_i \sim \text{Binomial}(n_i, \pi_i)$, where π_i is the probability of success in each trial, with $0 \leq \pi_i \leq 1$, where $\pi_i = y_i/n_i$. Thus, the probability of observing an outcome y_i can be written as

$$P(Y_i = y_i | n_i, y_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}; y_i = 0, 1, \cdots, n_i.$$

This probability depends on the number of known tests n_i, whereas the probability of success (π_i) is an unknown parameter. In Fig. 6.1, we observe that the probability of obtaining a normal plant depends on the applied dose of the demethylating agent. Given that y_i has a binomial distribution, the expected value (the mean) is the product of the number of trials and the probability of success in each trial, that is, $E(Y_i) = n_i \pi_i$. Since the number of trials is fixed (once the data have been obtained), modeling the probability of success is equivalent to modeling the expected value as well as the variance since it is also a function of the number of trials and the probability of success. So, the expected value and variance of y_i are

$$E(y_i) = \mu_i = n_i\pi_i; \text{Var}(y_i) = n_i\pi_i(1-\pi_i).$$

This variance is small if the value π_i is close to 0 or 1, and this increases to its maximum when $\pi_i = 0.5$. This can be seen in Fig. 6.1, where proportions close to 0 or 1 show less variance than do proportions between 0.1 and 0.2 for a demethylating agent dose of 0.5. This variance can also be written in terms of the expected value as:

$$\text{Var}(y_i) = \frac{\mu_i}{n_i}(n_i - \mu_i).$$

In this CRD, the fixed number of treatments t (doses) were randomly assigned to r experimental units (trays). The linear predictor describing the structure of the mean of this GLMM is

$$\eta_i = \eta + \tau_i$$

where η_i denotes the ith linear predictor, η is the intercept, and τ_i is the fixed effect due to treatments i ($i = 1, 2, \cdots, t$) with t treatments and r_i replicates in each treatment.

The components that define this GLMM are shown below:

Distribution: $y_i \sim \text{Binomial}(N_{ij}, \pi_i)$
Linear predictor: $\eta_i = \eta + \tau_i$
Link function: $\text{logit}(\pi_i) = \text{logit}\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i$

where η_i is the linear predictor that relates the effect of dose i ($i = 1, 2, \cdots, 6$) to probability π_i. The model uses the linear predictor (η_i) to estimate the means ($\pi_i = \mu_i$) of the observations for each treatment.

The following GLIMMIX program fits a CRD with a binomial response:

```
proc glimmix nobound method=Laplace;
class Dose Rep;
model y/N= dose/link=logit;
lsmeans dose/lines ilink;
run;
```

In this example, the distribution of the dataset was not specified to GLIMMIX in the model specification because by using the expression "`Y/N,`" proc GLIMMIX automatically infers that this dataset has a binomial distribution. It is also important to note that variable dose and repetition were declared as class variables in the "class" command, which Statistical Analysis Software (SAS) interprets as explanatory variables that are nonnumerical factors. However, the variable declared "Rep" is not used in the model specification.

Table 6.3 Results of the analysis of variance

(a) Fit statistics for conditional distribution	
−2 Log L (y | r. effects)	83.46
Pearson's chi-square	11.95
Pearson's chi-square/DF	0.50

(b) Type III tests of fixed effects

Effect	Num DF	Den DF	*F*-value	Pr > *F*
Dose	5	15	132.53	<0.0001

(c) Dose least squares (LS) means

Dose	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
0	3.9580	0.4122	15	9.60	<0.0001	0.9813	0.007581
0.01	3.9580	0.4122	15	9.60	<0.0001	0.9813	0.007581
0.1	2.0049	0.1728	15	11.60	<0.0001	0.8813	0.01808
0.5	−1.8360	0.1623	15	−11.31	<0.0001	0.1375	0.01925
1	−3.6633	0.3580	15	−10.23	<0.0001	0.02501	0.008729
1.5	−3.4337	0.3212	15	−10.69	<0.0001	0.03126	0.009728

Part of the results is shown in Table 6.3. Pearson's chi-squared statistic value divided by the degrees of freedom in part (a) (Pearson′s chi − square/DF = 0.5) indicates that there is no evidence of extra-dispersion in the dataset. The analysis of variance (ANOVA) tabulated in part (b) in Table 6.3, with the type III tests of fixed effects, indicates that there is a highly significant difference ($P = 0.0001$) in the average proportion of normal plants with respect to the dose applied to the seeds.

The output when using the "lsmeans" command in conjunction with the "ilink" option is in the "Mean" column (part (c) in Table 6.3). These values are the values of $\pi_i^{'}s$, i.e., the estimated probabilities $\hat{\pi}_0 = 0.9813$ and $\hat{\pi}_{0.01} = 0.9813$ of normal plants for the treatments whose doses are 0 and 0.01, respectively. For treatments with doses of 0.1 and 0.5, the observed probabilities of normal plants are $\hat{\pi}_{0.1} = 0.8813$ and $\hat{\pi}_{0.5} = 0.1375$, respectively, whereas for the 1 and 1.5 doses, the observed probabilities of normal plants decrease dramatically with $\hat{\pi}_1 = 0.02501$ and $\hat{\pi}_{1.5} = 0.03126$, respectively.

Figure 6.2 shows the mean comparisons (least significance difference (LSD)) of the estimated probabilities according to the dose applied to the seeds in trays. In this figure, we can observe that in the treatments with dose = 0 (control) and dose = 0.01, the observed proportions of normal plants are not statistically different from each other, but they do differ with the other applied doses. At a dose of 0.1, the observed proportion of normal plants was 88.13%, and this was statistically different from all the doses used. Finally, doses at 0.5, 1, and 1.5 of the demethylating agent in the observed proportion of normal plants decreased drastically to 13.75%, 2.501%, and 3.12%, respectively. The doses of 1 and 1.5 produced statistically equal proportions of normal plants.

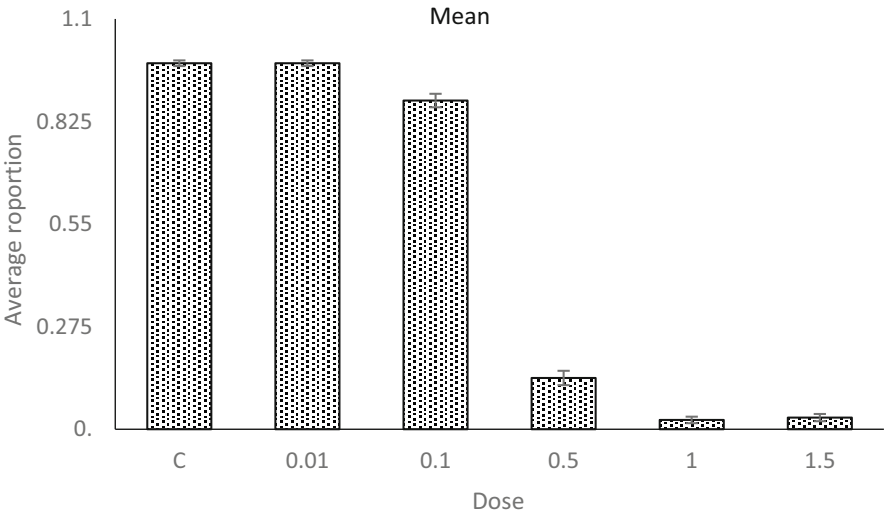

Fig. 6.2 Comparison of the estimated probabilities per dose of the demethylating agent

If the researcher wishes to model how dose levels of the demethylating agent affect normal plant proportions, then the dose must be declared as a continuous variable. The following SAS syntax with proc GLIMMIX runs a binomial regression:

```
proc glimmix data=crd_bin method=Laplace plots=all;
class Rep;
model y/N= dose/solution;
random rep;
run;quit.
```

Most of the commands and options have already been discussed throughout this book; the "model y/N" command indicates that the response variable is in a ratio. Therefore, this dataset is modeled with a binomial distribution, which is affected by the different number of individuals in each repetition. proc GLIMMIX interprets the distribution of the data as binomial, whereas the "solution" option requests the parameter estimates of the model (intercept and slope).

The components that define this GLMM are shown below:

Distribution: $y_i \sim \text{Binomial}(N_{ij}, \pi_i)$
Linear predictor: $\eta_i = \eta + \beta * \text{dose}_i$
Link function: $\text{logit}(\pi_i) = \text{logit}\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i$

Thus, the model can be written as

Table 6.4 Regression analysis results

(a) Fit statistics	
−2 Log likelihood	231.58
Akaike information criterion (AIC) (smaller is better)	235.58
AICC (smaller is better)	236.15
Bayesian information criterion (BIC) (smaller is better)	237.93
CAIC (smaller is better)	239.93
HQIC (smaller is better)	236.20
Pearson's chi-square	2317.12
Pearson's chi-square/DF	96.55

(b) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Dose	1	19	475.97	<0.0001

(c) Solutions for fixed effects					
Effect	Estimate	Standard error	DF	*t*-value	Pr > |*t*|
Intercept	2.7927	0.1302	3	21.46	0.0002
Dose	−7.6232	0.3494	19	−21.82	<0.0001

$$\eta_i = \log\left(\frac{\mu_i}{n_i - \mu_i}\right) = \log\left(\frac{n_i\pi_i}{n_i - n_i\pi_i}\right) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \text{logit}(\pi_i) = \eta + \beta\text{dose}_i$$

and the logit function can be written in terms of the probability of success, π_i, as

$$\pi_i = \frac{1}{1 + \exp(-\eta_i)}$$

Part of the SAS output of the GLIMMIX syntax is shown below. The goodness-of-fit statistics, type III tests of fixed effects, and parameter estimates are shown in Table 6.4. The analysis of variance indicates that the demethylating agent has a highly significant effect on the observed proportion of normal plants ($P < 0.0001$) (part (b)). The maximum likelihood estimates for the intercept and slope are $\eta = 2.7927$ and $\beta = -7.6232$, respectively.

Figure 6.3 shows that as the value of the linear predictor increases (η_i), the value of the residuals rapidly decreases. We can also see that the residuals plotted against the quantiles clearly do not follow a normal distribution because this model is not a linear function of the explanatory variable "dose."

Figure 6.4 shows that the proportions studied and fitted are not so far apart, and, as such, the binomial model is suitable for this dataset. The estimated linear predictor of this model is as follows:

$$\hat{\eta}_i = \hat{\eta} + \hat{\beta} \times \text{dose}_i = 2.7927 - 7.6232 \times \text{dose}_i.$$

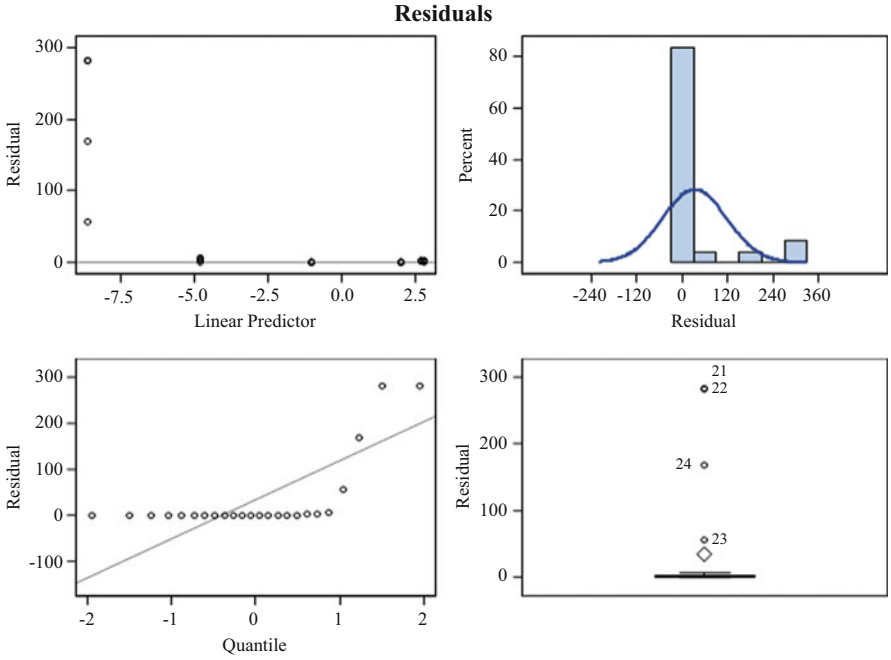

Fig. 6.3 A graph of residuals versus the linear predictor, quantiles

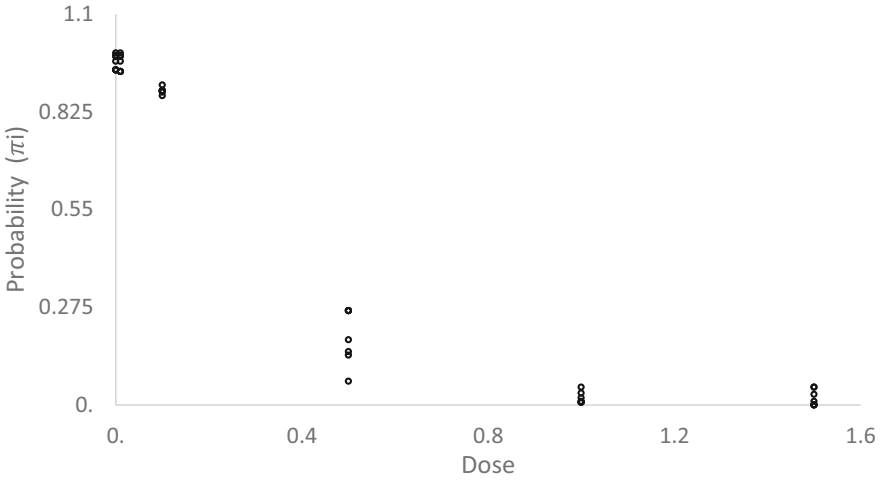

Fig. 6.4 Observed and estimated proportion

The logit of the probability of success is a linear function of the explanatory variables, so the model can be written in terms of the probability of success (observing normal plants) as

$$\pi_i = \frac{1}{1 + \exp^{(-\eta_i)}}$$

Given the parameter estimates, we can predict the success probability of observing a normal plant, and given a certain concentration of the demethylating agent, this estimated probability (using the estimated linear predictor) can be seen plotted in Fig. 6.4.

$$\hat{\pi}_i = \frac{1}{1 + \exp^{(\hat{\eta}_i)}} = \frac{1}{1 + \exp^{(-2.7927 + 7.6232 \times \text{dose}_i)}}$$

6.3 Factorial Design in a Randomized Complete Block Design (RCBD) with Binomial Data: Toxic Effect of Different Treatments on Two Species of Fleas

A group of researchers wishes to study the toxic effect of certain treatments (Trts) on two flea species (SP) (*Daphnia magna* and *Ceriodaphnia dubia*). To compare the toxicity effect of treatments on both flea species, a randomized complete block design (RCBD bioassay) was implemented with three replicates per treatment, with each replicate consisting of 10 fleas (Appendix: Fleas). The linear predictor describing this experiment is described below:

$$\eta_{ijkl} = \eta + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \text{bioassay}_k + \text{rep(bioassay)}_{l(k)}$$

where η is the intercept, α_i is the fixed effect due to species i, β_j is the fixed effect of treatment j, $(\alpha\beta)_{ij}$ is the fixed effects interaction between the flea species and treatment, bioassay_k is the random effect due to bioassay k assuming $\text{bioassay}_k \sim N\left(0, \sigma^2_{\text{bioassay}}\right)$, and $\text{rep(bioassay)}_{l(k)}$ is the random effect due to repetition bioassay assuming $\text{rep(bioassay)}_{l(k)} \sim N\left(0, \sigma^2_{\text{rep(bioassay)}}\right)$.

The remaining components of this GLMM with a binomial response (N_{ijk}, π_{ijk}) are described below:

Distribution: $y_{ijkl} \mid \text{bioassay}_k, \text{rep(bioassay)}_{l(k)} \sim \text{Binomial}(N_{ijk}, \pi_{ijk})$

$\text{bioassay}_k \sim N\left(0, \sigma^2_{\text{bioassay}}\right)$, $\text{rep(bioassay)}_{l(k)} \sim N\left(0, \sigma^2_{\text{rep(bioassay)}}\right)$, where N_{ijkl} is the number of dead fleas, observed in species i in replicate l in bioassay k under treatment j,

Link function: $\text{logit}\left(\pi_{ijk}\right) = \log\left[\frac{\pi_{ijk}}{1-\pi_{ijk}}\right] = \eta_{ijk}$.

The following SAS syntax allows us to fit the GLMM with a binomial response.

Table 6.5 Results of the analysis of variance

(a) Fit statistics	
−2 Log likelihood	145.33
AIC (smaller is better)	173.33
AICC (smaller is better)	177.85
BIC (smaller is better)	160.71
CAIC (smaller is better)	174.71
HQIC (smaller is better)	147.97

(b) Fit statistics for conditional distribution	
−2 Log L (Sobrevi | r. effects)	145.33
Pearson's chi-square	10.72
Pearson's chi-square/DF	0.10

(c) Covariance parameter estimates		
Cov Parm	Estimate	Standard error
Bioen	−0.1051	.
Bioen*SP (Rep)	−0.1192	.

(d) Type III tests of fixed effects				
Effect	Num DF	Den DF	F-value	Pr > F
SP	1	14	0.02	0.8829
Trt	5	80	15.08	<0.0001
SP*trt	5	80	4.66	0.0009

```
proc glimmix data=pulgas nobound method=laplace;
class Bioen SP Trt Rep ;
Model Sobrevi/n = SP|Trat/dist=binomial;
random Bioen sp*bioen(rep);
lsmeans SP|Trt/lines ilink;
run;
```

Part of the results is listed in Table 6.5. The fit statistics in part (a) and the conditional statistics in part (b) are useful for model comparison, whereas the variance component estimates are shown in part (c). The value of the statistic Pearson's chi − square/DF = 0.10 indicates that the binomial model gives a good fit to the dataset. The variance component estimates for bioassays and replication nested in bioassays are $\hat{\sigma}^2_{\text{bioassay}} = -0.1051$ and $\hat{\sigma}^2_{\text{rep(bioassay)}} = -0.1192$, respectively. The type III tests of fixed effects (part (d)) show the significance tests of the fixed effects in the model. The treatment effect and the interaction between the flea species (SP) and treatment are clearly significant with $P < 0.0001$ and $P = 0.0009$, respectively.

Since survival was statistically similar in both flea species, we will focus on the factors that were significant. Part (a) in Table 6.6 shows the means and standard errors of treatments on the model scale ("Estimate" column) and on the data scale ("Mean" column), obtained with "lsmeans" and the "ilink" option as well as the mean comparisons, which are on the model scale (part (b)).

Table 6.6 Means and standard errors on the model scale and on the data scale

(a) Trt least squares means

Trt	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
T1	8.1179	4.3180	80	1.88	0.0637	0.9997	0.001287
T2	4.3564	3.0554	80	1.43	0.1578	0.9873	0.03820
T3	1.0081	0.1924	80	5.24	<0.0001	0.7326	0.03768
T4	−1.0509	0.1712	80	−6.14	<0.0001	0.2591	0.03286
T5	−4.7187	3.0570	80	−1.54	0.1266	0.008848	0.02681
T6	−8.1182	4.3184	80	−1.88	0.0638	0.000298	0.001286

(b) Conservative T grouping of Trt least squares means (α=0.05)
LS means with the same letter are not significantly different

Trt	Estimate			
T1	8.1179		A	
T2	4.3564	B	A	
T3	1.0081	B	A	C
T4	−1.0509	B	D	C
T5	−4.7187		D	C
T6	−8.1182		D	

The LINES display does not reflect all significant comparisons. The following additional pairs are significantly different: (T3,T4)

Based on the fixed effects tests, the flea species × treatment interaction is significant. The means on the model scale are listed under the "Estimate" column, followed by their standard errors, "Standard error" (Table 6.7). The output of the "ilink" option in "lsmeans" applies the inverse function of the link function to the estimates on the model scale to obtain the estimates on the data scale. The probabilities, on the data scale, are given under the "Mean" column with their respective standard errors and correspond to the probability of insect (flea) survival.

Figure 6.5 shows that the survival of both species is different in treatments 2–5; the *Daphnia* species showed more resistance in treatments 2 and 3, whereas the *Ceriodaphnia* species showed greater resistance in treatments 4 and 5. On the other hand, in treatments 1 and 6, survival was similar in both species.

6.4 A Split-Plot Design in an RCBD with a Normal Response

A split plot is the most common treatment structure design in agricultural and agro-industrial research areas. These experiments generally involve two or more factors under study. Typically, large or primary experimental units, commonly known as the whole plot, are grouped into blocks. The levels of the first factor are randomly assigned to the whole plots. Then, each whole plot is divided into smaller units, known as split or secondary plots. The levels of the second factor are randomly assigned to the subplots within each whole plot.

Table 6.7 Means and standard errors on the model scale and on the data scale of the interaction between both factors

SP*treatment least squares means								
SP	Treatment	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
Daphnia	T1	8.1179	6.1065	80	1.33	0.1875	0.9997	0.001820
Daphnia	T2	8.1180	6.1068	80	1.33	0.1875	0.9997	0.001820
Daphnia	T3	1.9717	0.3218	80	6.13	<0.0001	0.8778	0.03452
Daphnia	T4	−1.2537	0.2536	80	−4.94	<0.0001	0.2221	0.04381
Daphnia	T5	−8.1186	6.1085	80	−1.33	0.1876	0.0002	0.001819
Daphnia	T6	−8.1182	6.1073	80	−1.33	0.1875	0.0002	0.001819
Ceriodaphnia	T1	8.1178	6.1064	80	1.33	0.1875	0.9997	0.001820
Ceriodaphnia	T2	0.5947	0.2202	80	2.70	0.0084	0.6444	0.05046
Ceriodaphnia	T3	0.04446	0.2109	80	0.21	0.8336	0.5111	0.05269
Ceriodaphnia	T4	−0.8480	0.2301	80	−3.69	0.0004	0.2999	0.04830
Ceriodaphnia	T5	−1.3188	0.2583	80	−5.11	<0.0001	0.2110	0.04301
Ceriodaphnia	T6	−8.1182	6.1071	80	−1.33	0.1875	0.0002	0.001819

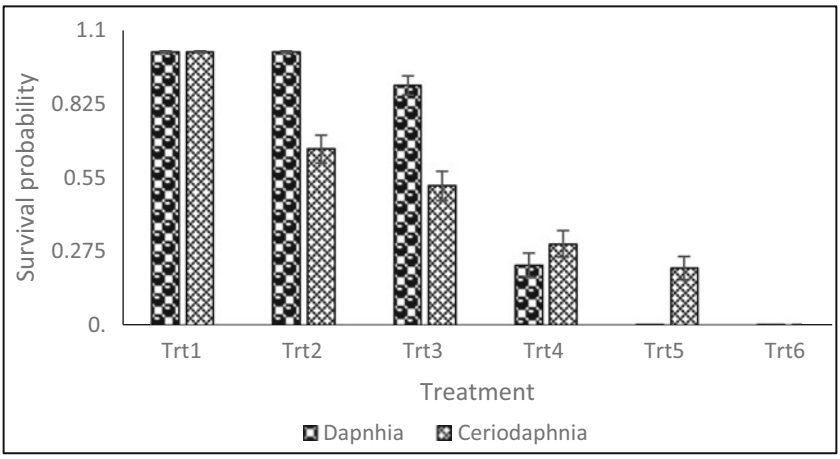

Fig. 6.5 The average survival rate of both species

The model equation for the analysis of variance assuming normality in the response is

$$y_{ijk} = \eta + \alpha_i + r_k + (\mathrm{ra})_{ik} + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

$$i = 1, 2, \cdots, a; j = 1, 2, \cdots, b; k = 1, 2, \cdots, r$$

where y_{ijk} is the observed response variable in the kth block at the ith level of factor A and at the jth level of factor B, α and β refer to the fixed treatment effects due to factors A and B, respectively, r is the random effect due to the blocks, $(ra)_{ik}$ is the random error term due to the whole plot that is an interaction between the blocks and factor A, and e_{ijk} is the random residual effect. Normally, the errors and other random terms are also assumed to be normal; however, when the response variable is not normally distributed, this way of specifying the model is not the most appropriate. Thus, under the assumption that the response variable is normal, this way of specifying the model is valid.

6.4.1 An RCBD Split Plot with Binomial Data: Carrot Fly Larval Infestation of Carrots

Data were obtained from an experiment that was designed to compare a number of carrot genotypes with respect to their resistance to infestation by carrot fly larvae. The data involved 16 genotypes that were compared at 2 pest levels to be controlled. The experiment was conducted in three randomized blocks. Each block consisted of

Table 6.8 The notation 44/53 denotes that 44 carrots were infected (y) out of a sample size of 53 studied (N)

	Treatment (level of infestation)					
	1			2		
Genotype	Block1	Block2	Block3	Block1	Block2	Block3
G1	44/53	42/48	27/51	16/60	9/52	26/54
G2	24/48	35/42	45/52	13/44	20/48	16/53
G3	8/49	16/49	16/50	4/52	6/51	12/43
G4	4/51	5/42	12/46	15/52	10/56	6/48
G5	11/52	13/51	15/44	4/51	6/43	9/46
G6	15/50	5/49	7/50	1/51	8/49	3/54
G7	18/52	13/47	7/47	2/52	4/52	6/52
G8	5/47	15/49	8/50	6/56	4/50	6/42
G9	11/52	6/45	5/51	3/54	8/51	3/53
G10	0/51	10/39	14/48	3/50	0/50	10/51
G11	6/52	4/46	10/37	1/52	7/38	4/48
G12	0/52	4/55	1/40	1/50	3/50	1/45
G13	14/45	18/43	4/40	4/51	7/46	7/45
G14	3/52	12/53	4/55	3/52	7/48	12/49
G15	11/52	6/54	5/49	2/50	4/46	14/53
G16	4/53	1/40	4/52	4/56	1/44	3/42

Table 6.9 Sources of variation and degrees of freedom

Sources of variation	Degrees of freedom
Blocks	$r - 1 = 3 - 1 = 2$
Factor A (infestation)	$a - 1 = 2 - 1 = 1$
Error_a (A*blocks)	$(r - 1)(a - 1) = 2$
Factor B (genotypes)	$b - 1 = 16 - 1 = 15$
Infestation*genotype (A*B)	$(a - 1)(b - 1) = 15$
Error_b	$a(r - 1)(b - 1) = 2 \times 2 \times 15 = 60$
Total	$r \times a \times b - 1 = 3 \times 2 \times 16 - 1 = 95$

32 plots, 1 for each combination of genotype and pest infestation level. At the end of the experiment, about 50 carrots were taken from each plot and assessed for infestation by carrot fly larvae. The data obtained are shown in Table 6.8.

Table 6.9 shows the analysis of variance summarizing the sources of variation and degrees of freedom.

Rewriting in terms of the linear predictor

$$\eta_{ijk} = \eta + \alpha_i + r_k + (\text{ra})_{ik} + \beta_j + (\alpha\beta)_{ij}$$

Since the observations were taken at the subplot level, conditioned on the structural effects of the design, these observations have a variance associated with the subplot. Therefore, α and β refer to the treatment fixed effects due to factors A

Table 6.10 Results of the analysis of variance

(a) Fit statistics for conditional distribution				
−2 Log L (y | r. effects)				527.82
Pearson's chi-square				189.09
Pearson's chi-square/DF				1.97
(b) Covariance parameter estimates				
Cov Parm	Subject	Estimate		Standard error
Intercept	Bloque	0.004272		0.02741
Trt	Bloque	0.03344		0.03545
(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Genotype	15	60	28.28	<0.0001
Trt	1	2	16.24	0.0564
Genotype*Trt	15	60	4.45	<0.0001

and B, respectively; $(\alpha\beta)_{ij}$ refers to the interaction of the above factors; r_k is the random effect due to blocks; and blocks × whole plot $(ra)_{ik}$ is assumed to contribute to the variation such that $r_k \sim N\left(0, \sigma_{\mathrm{r}}^2\right)$ and $(\mathrm{ra})_{ik} \sim N\left(0, \sigma_{\mathrm{block} \times A}^2\right)$. This model uses the linear predictor η_{ijk} to estimate the mean of the observations μ_{ijk}.

The specification of the this GLMM is as follows:

Distribution: $y_{ijk} \mid r_k, (ra)_{rk} \sim$Binomial$(N_{ijk}, \pi_{ijk})$
$r_k \sim N\left(0, \sigma_{\mathrm{r}}^2\right)$,
$(ra)_{rk} \sim N\left(0, \sigma_{\mathrm{block}*A}^2\right)$
Link function: $logit(\pi_{ijk}) = \eta_{ijk}$.

The following SAS GLIMMIX program allows the fitting of a GLMM with a split-plot structure in a randomized complete block design with a binomial response.

```
proc glimmix data=spd_pp nobound method=quadrature;
class Genotype Trt Block ;
model y/N = Genotype|Trt;
random intercept trt /subject=block;
lsmeans Genotype|Trt/lines ilink;
run;
```

The program uses the quadrature estimation method (**`method=quadrature`**). This estimation method produces similar results as the Laplace method. Part of the results is provided in Table 6.10. Pearson's chi-squared/DF value in part (a) gives an idea of whether there is overdispersion or extra-variation in the dataset. In this case, Pearson's chi − square/DF = 1.97 indicates that there is overdispersion in the dataset, so it is feasible to use either the pseudo-likelihood (PL) estimation method or a different distribution. In addition to these results, the variance component estimated due to blocks and blocks × genotype (the whole plot) in part (b) are $\sigma_{\mathrm{block}}^2 = 0.004272$ and $\sigma_{(\mathrm{block} \times A)}^2 = 0.03344$, respectively. The results of the fixed

effects tests (part (c)) indicate that the effect of genotype and the interaction between genotype and treatment are significant.

The appropriate method for model evaluation depends on whether or not there is evidence of overdispersion, so we consider this issue below. The residual variance incorporates systematic discrepancies between the model and the observed responses, variation between replicates (observations in independent experimental units with the same values of the explanatory variables) and sampling variation arising from the distribution of the data; in this case, it is the binomial distribution. If there are no duplicate observations and the fitted model provides an adequate description of the systematic trend, then only sampling variation contributes to the residual variance. If this is true, then the residual deviation has an approximate chi-squared distribution with degrees of freedom similar to the mean squared error (MSE) (the residual).

Since there is overdispersion in the data using the binomial distribution, there are three alternatives we can explore: (1) review the linear predictor, which involves carefully revising the analysis of variance table; (2) add a scale parameter; or (3) use another distribution for the dataset. Each of these three possible alternatives is discussed below, in this order.

6.4.1.1 Linear Predictor Review (η_{ijk})

If the proportion of normal plants (π_{ijk}) is being affected by the genotype within each infestation level (trt $= \alpha_i$) from plot to plot within each of the blocks, then a nested factorial effect of genotype within infestation levels (trt) could be included in the analysis of variance. Thus, the linear predictor would be defined as

$$\eta_{ijk} = \eta + \alpha_i + r_k + (\mathrm{ra})_{ik} + \beta(\alpha)_{j(i)}$$

where α_i, $\beta(\tau)_{j(i)}$, r_k, and $(ra)_{ik}$ are the fixed effects due to treatments, the effect of genotypes nested within a treatment, random effects due to blocks $\left(r_k \sim N\left(0, \sigma_{\mathrm{r}}^2\right)\right)$, and the interaction between blocks and treatment $\left((\mathrm{ra})_{ik} \sim N\left(0, \sigma_{\mathrm{RA}}^2\right)\right)$, respectively.

The following GLIMMIX syntax estimates the above linear predictor:

```
proc glimmix data=spd_pp method=laplace;
class Genotype Trt Block ;
model y/N = Trt genotype(trt);
random trt/subject=block;
lsmeans genotype(trt)/lines ilink slice=trt slicediff=trt;
run;
```

The only difference between this proc GLIMMIX and the previous one is that in this program, we have included the nested effect of genotypes within treatment, genotype (trt), and removed only the fixed effects of genotypes. Part of the results is shown in Table 6.11. The value of Pearson's chi-squared/DF statistic (part (a)) as

Table 6.11 Results of the analysis of variance, under a new linear predictor

(a) Fit statistics for conditional distribution				
−2 Log L (y | r. effects)				527.82
Pearson's chi-square				189.07
Pearson's chi-square/DF				1.97
(b) Covariance parameter estimates				
Cov Parm	Subject	Estimate		Standard error
Intercept	Bloque	0.004265		0.02740
Trt	Bloque	0.03343		0.03544
(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Trt	1	2	16.20	0.0565
Genotype (Trt)	30	60	15.83	<0.0001

well as the fit statistics did not decrease when modifying the linear predictor. However, the *F*-values calculated for treatments and genotypes within treatments (part (c)) are smaller than those obtained in the split-plot design.

Since the overdispersion is still present (Pearson's chi − square/DF = 1.97), another alternative is to add a scaling parameter to the model. This alternative is presented below.

6.4.1.2 Scale Parameter

If the residual deviation is larger than expected when compared to critical values of the appropriate chi-squared distribution, and if this cannot be corrected by redefining the linear predictor of the model, then there is more variation present than can be accounted for by the distributional likelihood assumption. In this case, we say that the data show overdispersion. The simplest way to deal with overdispersion is to extend the model for scaling the variance function. Adding the scale parameter replaces $\text{Var}(y_{ij}) = \pi_{ij}(1 - \pi_{ij})$ with $\text{Var}(y_{ij}) = \phi\pi_{ij}(1 - \pi_{ij})$. The rationale for this approach is discussed by Collett (2002). The parameter ϕ is a scale factor, called the dispersion parameter, which is used to summarize the degree of overdispersion present in the observations. Clearly, $\phi = 1$ corresponds to the original distribution model. This parameter can be estimated in several different ways. The logarithm of the likelihood of the binomial distribution is given by

$$\log\binom{N}{y_{ij}} + y_{ij}\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) + N\log\left(1-\pi_{ij}\right)$$

In the logarithm of the likelihood, the term "$y_{ij}\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$" is very important; any quantity that multiplies y_{ij} is known as the natural or canonical parameter, and this parameter is always a function of the mean. For the binomial distribution, the mean

$N_{ij}\pi_{ij}$ and the natural parameter is $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$, and, in categorical data, it is known as "log odds." The generalized estimating equation (GEE) method provides a valid analysis for marginal means, since under a binomial distribution, in the quasi-likelihood, the variance of the distribution is given by $\phi\pi_{ij}(1-\pi_{ij})$. This is achieved by adding the "random _residual_" command in the following SAS syntax.

The following GLIMMIX commands are used to invoke the scale parameter but using the first predictor proposed for these data.

```
proc glimmix data=spd_pp nobound;
class GenotypeTrtBlock ;
model y/N = Trt|genotype;
random intercept trt/subject=block;
random _residual__;
lsmeans Trt|genotype/lines ilink ;
run;
```

In this syntax, we still keep the binomial distribution (y/N is equivalent to telling GLIMMIX in SAS that it is a binomial response) but will add the "random _residual_" command. In this case, we cannot obtain the maximum likelihood estimators because we cannot implement the Laplace method ("method = laplace") or adaptive quadrature ("method = quad") approximation method, so the estimation is performed through the pseudo-likelihood (PL) method. This causes the scale parameter to be estimated, and, consequently, it is used in the adjustment of all standard errors and statistical tests. Proc GLIMMIX uses the generalized statistics of McCullagh and Nelder (1989), i.e., χ^2/df as the estimator of the scale parameter ($\hat{\phi}$). All standard errors from the analysis under a binomial distribution are multiplied by $\sqrt{\hat{\phi}}$, and all F-tests are divided by $\hat{\phi}$ to account for overdispersion. Part of the output is shown below.

The value of Pearson's statistic in part (a) indicates that overdispersion has not been eliminated. Chi − square/DF = 3.13, on the contrary, indicates that this value has increased. This result indicates that adding a scale parameter to the model does not decrease the extra-variation present in the dataset, since the binomial assumption forces a relationship between the mean and variance of the data that might not contain the data being analyzed. On the other hand, the estimated scale parameter is $\hat{\phi} = 3.1263$ (part (b)). Pearson's residual analysis showed that its variance is 3.6257, which is considerably larger than 1, implying a large overdispersion. In addition, the results of the fixed effects tests (part (c)) vary from those above (Table 6.12).

Therefore, the third option based on assuming an alternative distribution (beta distribution) on the response variable is discussed below.

Table 6.12 Results of the analysis of variance, adding a scale parameter to the model

(a) Fit statistics	
−2 Res log pseudo-likelihood	182.52
Generalized chi-square	200.09
Gener. chi-square/DF	3.13

(b) Covariance parameter estimates			
Cov Parm	Subject	Estimate	Standard error
Intercept	Bloque	0.005416	0.04750
Trt	Bloque	0.03202	0.06338
Residual variance component (VC)		3.1263	0.5719

(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	F-value	Pr > F
Trt	1	2	10.20	0.0856
Genotype	15	60	9.04	<0.0001
Genotype*Trt	15	60	1.42	0.1674

6.4.1.3 Alternative Distribution

Another approach to control the overdispersion would be to use a different distribution in the interval [0, 1], such as the beta distribution, to model the data. Generally, this distribution yields good results when all experiments have the same number of observations (successes and failures), i.e., when $N_{ijk} = N$. When N_{ijk} varies a little, even in many cases, the beta distribution yields acceptable results. It is important to mention that the proportions come from binomial counts, and, therefore, we now define the response variable as $p_{ijk} = \frac{y_{ijk}}{N_{ijk}}$ so that it can be modeled as the beta distribution. The components of the beta response model are listed below:

Distribution: $p_{ijk} \mid r_k, (ra)_{rk} \sim \text{Beta}(\pi_{ijk}, \phi)$ with ϕ as the scale parameter
$r_k \sim N\left(0, \sigma_{\text{r}}^2\right)$, $(\text{ra})_{rk} \sim N\left(0, \sigma_{\text{RA}}^2\right)$
Linear predictor: $\eta_{ijk} = \eta + \alpha_i + r_k + (\alpha\text{r})_{ik} + \beta_j + (\alpha\beta)_{ij}$
Link function: $\text{logit}\left(\pi_{ijk}\right) = \text{logit}\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \eta_{ijk}$

As mentioned before, we now use the response variable $p_{ijk} = \frac{y_{ijk}}{N_{ijk}}$. This new response variable p_{ijk} is not the same as the one used in the binomial distribution. The following SAS commands fit a GLMM in a split-plot randomized complete block design with a beta response. It is important to mention that before implementing this model in SAS GLIMMIX, the variable $p = p_{ijk} = \frac{y_{ijk}}{N_{ijk}}$ was defined.

```
proc glimmix data=spd_pp nobound method=laplace;
class GenotypeTrtBlock ;
model p = Genotype|Trt/dist=beta;
random intercept trt/subject=block;
lsmeans Genotype|Trt/lines ilink;
run;
```

Table 6.13 Fit statistics assuming binomial and beta distributions

(a) Fit statistics		
Distribution	Binomial	Beta
−2 Log likelihood	541.85	−246.49
AIC (smaller is better)	609.85	−176.49
AICC (smaller is better)	648.87	−132.28
BIC (smaller is better)	579.20	−208.04
CAIC (smaller is better)	613.20	−173.04
HQIC (smaller is better)	548.24	−239.91
(b) Fit statistics for conditional distribution		
Distribution	Binomial	Beta
−2 Log L (y | r. effects)	527.82	−254.68
Pearson's chi-square	189.09	93.95
Pearson's chi-square/DF	1.97	1.01

Table 6.14 Results of the analysis of variance, assuming binomial and beta distributions

(a) Covariance parameter estimates					
		Binomial		Beta	
Cov Parm	Subject	Estimate	Standard error	Estimate	Standard error
Intercept	Bloque	0.004272	0.02741	−0.00524	.
Trt	Bloque	0.03344	0.03545	0.02175	0.1475
Scale $(\hat{\phi})$			.	25.7070	

(b) Type III tests of fixed effects						
			Binomial		Beta	
Effect	Num DF	Den DF	F-value	Pr > F	F-value	Pr > F
Trt	1	4	16.24	0.0564	9.98	0.0342
Genotype	15	60	28.28	<0.0001	13.25	<0.0001
Genotype*Trt	15	60	4.45	<0.0001	2.23	0.0146

Some of the SAS GLIMMIX output is listed below. Based on the fit statistics under the binomial (first alternative) and beta distributions (Table 6.13), clearly the values of the statistics related to the degree of overdispersion are lower in the beta distribution than in the binomial distribution, indicating that the beta distribution provides a better fit (part (a)). Looking at the fit statistics for the conditional model in part (b), the values of the three fit statistics in the binomial model are higher than the values in the beta model. The value of Pearson's chi − square/DF under the beta distribution is 1.01. This value indicates that the overdispersion has been virtually eliminated from the data and that therefore the beta distribution is a better candidate model for this dataset.

Adding the scale parameter (ϕ) to the model, the variance components and standard errors in Table 6.14 cause (part (a)) variation for each of the results and, therefore, the F- and t-tests are affected (part (b)). The estimated value of the scale

Table 6.15 Estimated means and standard errors on the model scale and the data scale

(a) Trt least squares means

Trt	Estimate	Standard error	DF	*t*-value	Pr > |t|	Mean	Standard error mean
Trt1	−1.2362	0.01768	2	−69.94	0.0002	0.2251	0.003083
Trt2	−1.9327	0.01768	2	−109.34	<0.0001	0.1264	0.001952

(b) Genotype least squares means

Genotype	Estimate	Standard error	DF	*t*-value	Pr > |t|	Mean	Standard error mean
G1	0.1524	0	57	Infty	<0.0001	0.5380	0
G10	−1.4143	0	57	−Infty	<0.0001	0.1956	0
G11	−1.8698	0	57	−Infty	<0.0001	0.1336	0
G12	−2.8971	0.03885	57	−74.58	<0.0001	0.05230	0.001925
G13	−1.4336	0	57	−Infty	<0.0001	0.1925	0
G14	−1.8761	0.1304	57	−14.39	<0.0001	0.1328	0.01502
G15	−1.8618	0	57	−Infty	<0.0001	0.1345	0
G16	−2.6686	0	57	−Infty	<0.0001	0.06485	0
G2	0.2225	0	57	Infty	<0.0001	0.5554	0
G3	−1.3329	0	57	−Infty	<0.0001	0.2087	0
G4	−1.5897	0	57	−Infty	<0.0001	0.1694	0
G5	−1.3696	0	57	−Infty	<0.0001	0.2027	0
G6	−2.0173	0	57	−Infty	<0.0001	0.1174	0
G7	−1.7001	0.1356	57	−12.53	<0.0001	0.1545	0.01771
G8	−1.7161	0	57	−Infty	<0.0001	0.1524	0
G9	−1.9796	0	57	−Infty	<0.0001	0.1214	0

parameter is $\hat{\phi} = 25.7018$. The variance components based on the binomial model and beta are listed below.

The treatment means (part (a)) and genotypes (part (b)) are presented in Table 6.15. The estimates on the model scale are listed under the column "Estimate" with their respective standard errors "Standard error," and the values on the data scale are listed under the column "MEAN" with their respective standard errors "Standard error mean." In the table of least squares means for the effect of genotypes, inconsistencies are observed in the values of *t* and in the standard error values of the means, so other estimation alternatives should be sought.

In large samples, both binomial and normal distributions are quite similar. Logically, the latter two analyses, binomial and beta, are attractive because of their consistency with the nature of the data. Because of the inconsistencies in the estimates of the mean for genotypes (tvalue = Infty and standard error of the mean), a robust method of estimation could be used; in this case, this is the normal distribution.

Assuming that p_{ijk} has a normal distribution with a mean μ_{ijk} and constant variance σ^2, the components of this model are as follows:

Table 6.16 Results of the analysis of variance, assuming a normal distribution

(a) Fit statistics	
−2 Res log likelihood	−79.38
AIC (smaller is better)	−73.38
AICC (smaller is better)	−72.98
BIC (smaller is better)	−76.08
CAIC (smaller is better)	−73.08
HQIC (smaller is better)	−78.81
Generalized chi-square	0.60
Gener. chi-square/DF	0.01

(b) Covariance parameter estimates		
Cov Parm	Estimate	Standard error
Bloque	0.000123	0.000742
Trt*bloque	0.000329	0.000925
Residual	0.009442	0.001724

(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	F-value	Pr > F
Genotype	15	60	12.59	<0.0001
Trt	1	2	20.46	0.0456
Genotype*Trt	15	60	2.93	0.0016

Distribution: $\text{pct}_{ijk} \mid r_k, (\text{ra})_{ik} \sim \text{Normal}(\mu_{ijk}, \sigma^2)$
$r_k \sim N(0, \sigma_r^2)$, $(\text{ra})_{ik} \sim N(0, \sigma_{\text{RA}}^2)$
Linear predictor: $\eta_{ijk} = \eta + \alpha_i + r_k + (\alpha r)_{ik} + \beta_j + (\alpha\beta)_{ij}$
Link function: $\eta_{ijk} = \mu_{ijk}$; identity

Similarly, in this example, the response variable used was $\text{pct}_{ijk} = \frac{y_{ijk}}{N_{ijk}}$. This new response variable pct_{ijk} is not the same as the response variable used in the binomial distribution. The following SAS GLIMMIX commands adjust a linear mixed model (LMM) under a split plot in a randomized complete block design with a normal response.

```
proc glimmix data=spd_pct nobound;
class Genotype Trt Block ;
model pct = Genotype|Trt;
random block block*trt;
lsmeans Genotype|Trt/lines;
run;
```

Part of the results is shown below. The values of fit statistics in part (a) of Table 6.16 for the model are clearly lower than those estimated in the previous options. This indicates that the normal distribution is reasonable, even though the response is a proportion. The estimated variance components, tabulated in part (b) due to blocks, blocks x treatment, and the mean squared error (MSE) (Residual = Gener. chi-square/DF) are $\hat{\sigma}^2_{\text{block}} = 0.000123$, $\hat{\sigma}^2_{\text{block} \times \text{trt}} = 0.00039$, and $\hat{\sigma}^2 = \text{MSE} = 0.009442 \cong 0.01$, respectively.

Table 6.17 Means and standard errors for genotypes and treatments

(a) Genotype least squares means

Genotype	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
G1	0.5260	0.04086	60	12.87	<0.0001	0.5260	0.04086
G10	0.1340	0.04086	60	3.28	0.0017	0.1340	0.04086
G11	0.1522	0.04086	60	3.73	0.0004	0.1522	0.04086
G12	0.03332	0.04086	60	0.82	0.4179	0.0333	0.04086
G13	0.2026	0.04086	60	4.96	<0.0001	0.2026	0.04086
G14	0.1342	0.04086	60	3.28	0.0017	0.1342	0.04086
G15	0.1360	0.04086	60	3.33	0.0015	0.1360	0.04086
G16	0.05625	0.04086	60	1.38	0.1737	0.0562	0.04086
G2	0.5355	0.04086	60	13.11	<0.0001	0.5355	0.04086
G3	0.2139	0.04086	60	5.24	<0.0001	0.2139	0.04086
G4	0.1751	0.04086	60	4.28	<0.0001	0.1751	0.04086
G5	0.2035	0.04086	60	4.98	<0.0001	0.2035	0.04086
G6	0.1301	0.04086	60	3.18	0.0023	0.1301	0.04086
G7	0.1671	0.04086	60	4.09	0.0001	0.1671	0.04086
G8	0.1504	0.04086	60	3.68	0.0005	0.1504	0.04086
G9	0.1187	0.04086	60	2.90	0.0051	0.1187	0.04086

(b) Trt least squares means

Trt	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
Trt1	0.2478	0.01863	2	13.30	0.0056	0.2478	0.01863
Trt2	0.1358	0.01863	2	7.29	0.0183	0.1358	0.01863

The *F*-statistics for the fixed effects of genotype, treatments, and the interaction between both factors provide significant statistical evidence on the proportion of infested carrots in each of the genotypes (part (c)). Overall, the least squares means for genotypes and treatments are reported in Table 6.17 in parts (a) and (b). The genotypes showing the highest fraction of infested carrots were 1, 2, 3, 5, and 13, whereas genotypes 12 and 16 showed the lowest percentage of infested carrots. Now, for treatments, the highest proportion of infested carrots was observed in treatment 1 with 24.78%, whereas in treatment 2, it was 13.58%.

Based on the fixed effects tests, the interaction effect of genotype x treatment on the proportion of infested carrots was statistically different. Genotypes 9 and 16 showed higher susceptibility in treatment 1 followed by treatment 2, whereas genotypes 5, 11, 13, and 15 showed the same proportions of infested carrots in both treatments (Fig. 6.6). On the other hand, genotypes that showed higher resistance to infestation levels were genotypes 1, 2, and 6 followed by genotypes 3, 4, 7, 8, 10, and 12.

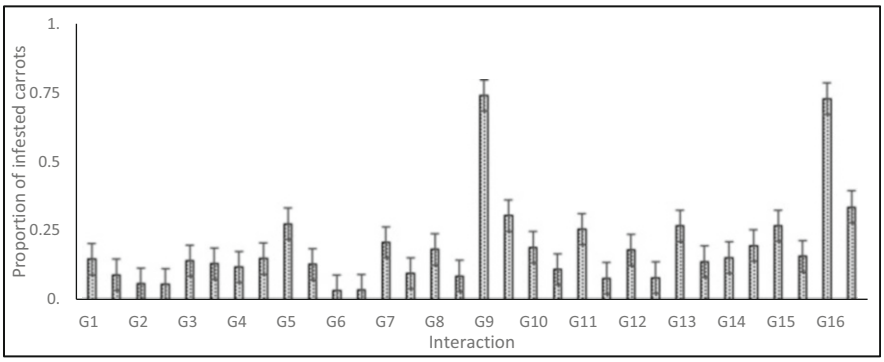

Fig. 6.6 The average proportion of infested carrots in genotypes as a function of treatment

6.5 A Split-Split Plot in an RCBD:- In Vitro Germination of Seeds

The growth of a plant in a tissue culture can be explained by various combined effects of A, B, and C factors. For this, the availability and efficient use of chemical resources (factors) is of great relevance when availability is scarce or too expensive. In light of this, the combination of three reagents (A, B, and C), reagent A at three levels and reagents B and C at two levels, were tested on the in vitro germination of orchid seeds. The combination of the levels of each of the factors is schematized below.

Block 1											
A_3				A_1				A_2			
B_1		B_2		B_1		B_2		B_2		B_1	
C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1

Block 2											
A_2				A_1				A_3			
B_1		B_2		B_1		B_2		B_2		B_1	
C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_1
C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2

In each of the factor combinations, N orchid seeds were placed to germinate for a period of time. Let y_{ijk} be the number of seeds germinated at the ith level of factor A, at the jth level of factor B, and at the kth level of factor C. Since the observations are made at the sub-subplot level, conditional on the structural effects of the design, these observations have a variance associated with the subplot. Therefore, the statistical model for this experiment is given below:

Table 6.18 Number of seeds that germinated (y_{ijkl}) in each of the factor combinations

Block	A	B	C	Y	N
1	1	1	1	15	73
2	1	1	1	10	86
1	1	1	2	17	69
2	1	1	2	19	32
1	1	2	1	26	125
2	1	2	1	21	62
1	1	2	2	14	81
2	1	2	2	12	21
1	2	1	1	10	92
2	2	1	1	12	108
1	2	1	2	30	44
2	2	1	2	32	33
1	2	2	1	37	91
2	2	2	1	30	42
1	2	2	2	32	98
2	2	2	2	37	44
1	3	1	1	18	52
2	3	1	1	18	73
1	3	1	2	23	108
2	3	1	2	21	55
1	3	2	1	24	106
2	3	2	1	27	92
1	3	2	2	37	64
2	3	2	2	37	97

Distribution: $y_{ijkl} \mid r_l, (ra)_{il}, (r\alpha\beta)_{ijl} \sim \text{Binomial}(N_{ijk}, \pi_{ijk})$

$r_l \sim N\left(0, \sigma_r^2\right)$, $(\text{ra})_{rk} \sim N\left(0, \sigma_{\text{RA}}^2\right)$, $(r\alpha\beta)_{ijl} \sim N\left(0, \sigma_{\text{rab}}^2\right)$

Linear predictor:

$\eta_{ijk} = \eta + \alpha_i + r_l + (r\alpha)_{il} + \beta_j + (\alpha\beta)_{ij} + (r\alpha\beta)_{ijl} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$,

where blocks (r_l), blocks × A ($(\text{ra})_{il}$), and blocks × A × B ($(r\alpha\beta)_{ijl}$) are assumed to contribute to the variation such that $r_l \sim N\left(0, \sigma_{\text{r}}^2\right)$, $(\text{ra})_{il} \sim N\left(0, \sigma_{\text{r}\times\text{A}}^2\right)$, $(r\alpha\beta)_{ijl} \sim N\left(0, \sigma_{\text{rab}}^2\right)$, respectively, and ε_{ijkl} experimental errors are distributed as $N(0, \sigma^2)$. This model uses the linear predictor η_{ijk} to estimate the mean of the observations μ_{ijk}.

Link function: $\text{logit}(\pi_{ijkl}) = \eta_{ijkl}$

Table 6.18 below shows the data obtained from this experiment.

Table 6.19 presents the analysis of variance and shows the sources of variation and degrees of freedom for this experimental design.

The following SAS GLIMMIX program allows a GLMM with a split-split plot structure to be fitted in an RCBD with a binomial response.

Table 6.19 Sources of variation and degrees of freedom for the randomized block design with an arrangement of treatments under the split-split-plot structure

Sources of variation	Degrees of freedom
Blocks	$r - 1 = 2 - 1 = 1$
Factor A	$a - 1 = 3 - 1 = 2$
Error_a(Bloque*A)	$(r - 1)(a - 1) = 2$
Factor B	$b - 1 = 2 - 1 = 1$
A* B	$(a - 1)(b - 1) = 2$
Error_b(A*B(Bloque))	$a(b - 1)(r - 1) = 3 \times 1 \times 1 = 3$
Factor C	$(c - 1) = 2 - 1 = 1$
A*C	$(3 - 1)(2 - 1) = 2$
B*C	$(b - 1)(c - 1) = 1$
A*B* C	$(a - 1)(b - 1)(c - 1) = 2$
Error	$ab(c - 1)(r - 1) = 3 \times 2 \times 1 \times 1 = 6$
Total	$r \times a \times b \times c - 1 = 2 \times 3 \times 2 \times 2 - 1 = 23$

```
proc GLIMMIX data=germ nobound method=laplace;
class Block A B C;
model Y/N = A|B|C/dist=binomial link=logit;
random block block*A block*A block*A*B;
lsmeans A|B|C/lines ilink;
run;
```

Part of the output is shown in Table 6.20. The value of the conditional statistic Pearson' chi − square/DF = 1.81 (part (a)) indicates that there is an overdispersion in the dataset since these values are greater than 1. The estimated variance components tabulated in part (b) correspond to blocks, blocks × factor A, and blocks × factor A × factor B, which are $\sigma^2_{\text{r}} = 0.0752, \sigma^2_{\text{rA}} = 0.088$, and $\sigma^2_{\text{rab}} = 0.0425$, respectively. The type III tests of fixed effects are shown in part (c). Here, we see that the test of equality of treatments is not significant for factors A and B and the interaction AB ($A, P = 0.1917, B, P = 0.0897$; AB, $P = 0.6262$), whereas for factor C and the interactions AC, BC, and ABC, it is significant at a level of 5%.

Since there is overdispersion in the dataset, the binomial distribution does not provide a good fit for the dataset (Pearson's chi − square/DF = 1.81). An alternative to model this dataset could be the beta distribution. Under this assumption, let the response variable be $p_{ijk} = \frac{y_{ijk}}{N_{ijk}}$, the proportion of seeds that germinated, then p_{ijk} is assumed to have a beta distribution rather than a binomial distribution for the success count y_{ijk} out of a total of N_{ijk} Bernoulli trials.

The components of the model are listed below:

Distribution: $p_{ijk} \mid r_l, (ra)_{il}, (r\alpha\beta)_{ijl} \sim Beta(\pi_{ijk}, \phi)$, with ϕ as the scale parameter.
$r_l \sim N\left(0, \sigma^2_r\right), (ra)_{rk} \sim N\left(0, \sigma^2_{rA}\right), (r\alpha\beta)_{ijl} \sim N\left(0, \sigma^2_{rab}\right)$
Linear predictor:
$\eta_{ijk} = \eta + \alpha_i + r_l + (r\alpha)_{il} + \beta_j + (\alpha\beta)_{ij} + (r\alpha\beta)_{ijl} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$
Link function: $\text{logit}\left(\pi_{ijk}\right) = \text{logit}\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \eta_{ijk}$

Table 6.20 Results of the analysis of variance of the RCBD in the split-split plot under the binomial distribution

(a) Fit statistics for conditional distribution	
−2 Log L (y | r. effects)	146.19
Pearson's chi-square	43.49
Pearson's chi-square/DF	1.81

(b) Covariance parameter estimates		
Cov Parm	Estimate	Standard error
Bloque	0.07521	0.1180
Bloque*A	0.08847	0.09319
Bloque*A*B	0.02205	0.04258

(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
A	2	2	4.22	0.1917
B	1	3	6.12	0.0897
A*B	2	3	0.55	0.6262
C	1	6	65.73	0.0002
A*C	2	6	11.68	0.0085
B*C	1	6	29.38	0.0016
A*B*C	2	6	31.69	0.0006

The following SAS commands fit a GLMM on a split-split plot in a randomized complete block design assuming a beta distribution for the response variable.

```
proc glimmix data=germ nobound method=laplace;
class BlockABC ;
model p = A|B|C/dist=beta ;
random block block*A block*A*B;/*intercept A /subject=block*/;
lsmeans A|B|C/lines ilink;
run;
```

Part of the results is listed in Table 6.21 under a beta distribution. The value of the fit statistic for the conditional model tabulated in (a) (Pearson's chi − square/DF = 1.01) indicates that overdispersion has been removed and that the beta distribution is a good model to fit the dataset. Part (b) shows the variance component estimates for blocks, *blockxA*, and *blockxAxB* $\left(\hat{\sigma}^2_{\mathrm{r}} = -0.157, \sigma^2_{\mathrm{rA}} = -0.05558, \text{and } \sigma^2_{\mathrm{rab}} = -0.227, \text{respectively}\right)$ and the value of the estimated scale parameter $\left(\hat{\phi} = 19.2789\right)$. According to the type III tests of fixed effects in part (c), the main effect of factor C ($P = 0.0128$) and interaction A×B×C ($P = 0.0424$) are statistically significant at a level of 5%.

The estimates of the interactions are shown in Table 6.22 on the model scale under the "Estimate" column and as probabilities on the data scale under the "Mean" column with its corresponding standard errors under the "Standard error mean" column.

Table 6.21 Results of the analysis of variance of the RCBD in the split-split plot structure under the beta distribution

(a) Fit statistics for conditional distribution	
−2 Log L (p | r. effects)	−37.51
Pearson's chi-square	21.31
Pearson's chi-square/DF	1.01

(b) Covariance parameter estimates		
Cov Parm	Estimate	Standard error
Bloque	−0.1570	.
Bloque*A	−0.05558	.
Bloque*A*B	−0.2270	.
Scale	19.2789	5.8703

(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
A	2	2	1.21	0.4521
B	1	2	0.00	0.9687
A*B	2	2	1.08	0.4799
C	1	4	18.34	0.0128
A*C	2	4	1.50	0.3257
B*C	1	4	6.56	0.0626
A*B*C	2	4	7.72	0.0424

Table 6.22 Estimated least mean squares on the model scale ("Estimate" column) and the data scale ("Mean" column)

A*B*C least squares means									
A	B	C	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
1	1	1	−0.3769	0.3194	4	−1.18	0.3034	0.4069	0.07709
1	1	2	0.9506	0.3445	4	2.76	0.0509	0.7212	0.06927
1	2	1	0.1721	0.3147	4	0.55	0.6135	0.5429	0.07810
1	2	2	0.7010	0.3308	4	2.12	0.1014	0.6684	0.07331
2	1	1	−0.6521	0.3296	4	−1.98	0.1190	0.3425	0.07422
2	1	2	2.9148	0.8071	4	3.61	0.0225	0.9486	0.03937
2	2	1	0.7430	0.4699	4	1.58	0.1890	0.6776	0.1026
2	2	2	0.4056	0.4515	4	0.90	0.4198	0.6000	0.1084
3	1	1	0.2695	0.3161	4	0.85	0.4419	0.5670	0.07761
3	1	2	0.2752	0.3163	4	0.87	0.4334	0.5684	0.07759
3	2	1	0.1236	0.3143	4	0.39	0.7143	0.5309	0.07827
3	2	2	1.1726	0.3614	4	3.24	0.0315	0.7636	0.06523

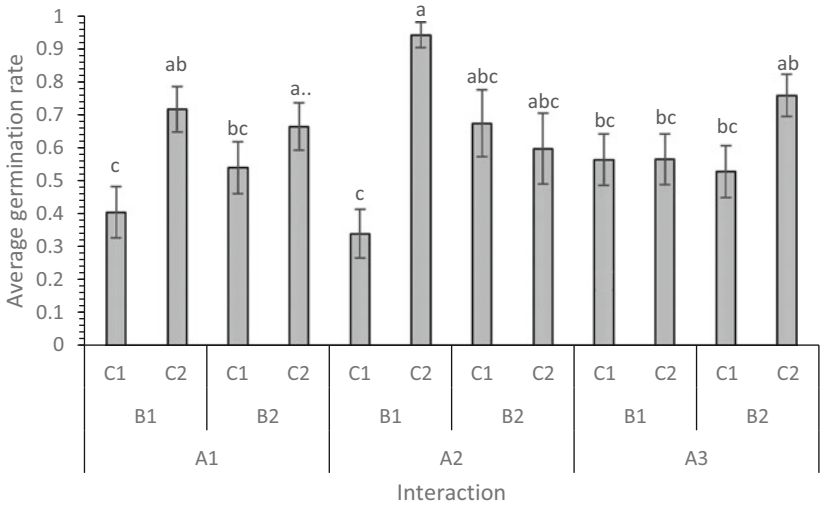

Fig. 6.7 The average seed germination rate

The simple effects of factors show that the best combination of factor levels was $A2{*}B1{*}C2$, showing the highest seed germination proportion followed by the combination of factors $A1{*}B1{*}C2$, $A3{*}B2{*}C2$, and lower proportion, which were observed in the combination of factors $A1{*}B2{*}C2$, $A2{*}B2{*}C1$ and $A2{*}B2{*}C2$ (Fig. 6.7). Finally, the combination of the factor levels $A2 \times B1 \times C1$ showed the lowest proportion of seed germination.

6.6 Alternative Link Functions for Binomial Data

In previous chapters, we used proc GLIMMIX with binomial data and, by default, it works with the link function "logit." However, in certain applications with binomial data, other link functions are acceptable, either because they make it easier to interpret or because for certain binomial datasets, the link function "logit" cannot accurately model the data and, as a result, produce biased (misleading) results. In this section, we consider two alternative link functions to the logit for binomial data: the link "probit" and the complementary log-log link.

The probit model is also used to model dichotomous (Bernoulli) or binomial (sum of Bernoulli trials) responses. For this model, the link function, called the probit link, uses the inverse of the cumulative distribution function of a standard normal distribution to transform probabilities to the standard normal variable. That is, $\Phi^{-1}(\pi_i) = \eta_i$, which implies that $\pi_i = \Phi(\eta_i)$, where $\Phi(Z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$.

The use of the probit regression model dates back to Bliss (1934). Bliss was interested in finding an effective pesticide to control insects that fed on grape leaves.

He discovered that the relationship between the response and a dose of pesticide was sigmoid, and he applied the probit link function to transform the dose–response curve from a sigmoid to a linear relationship.

The complementary function log − log defined as $\eta_i = \log(-\log(1-\pi_i))$, whose inverse is $\pi_i = 1 - e^{-e^{\eta_i}}$, is useful for data in which most of the probabilities are near zero or near one. For small values of π_i, the log-log transformation produces results highly similar to those produced when using a logit link. As the probability increases, the transformation approaches infinity more slowly than the probit or logit model.

6.6.1 Probit Link: A Split-Split Plot in an RCBD with a Binomial Response

This example takes the dataset of the split-split plot in an RCBD (Exercise 6.8.5). In this example, the data were modeled using the function "logit." In this exercise, we will fit the dataset using the link function "probit," and we will compare and contrast the results using a logit link. The components of the GLMM are identical to those in Example 6.5, except for the link function. That is, we replace:

$$\text{Link function: } \text{logit}(\pi_{ijk}) = \text{logit}\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \eta_{ijk} \text{ by } \Phi^{-1}(\pi_{ijk}) = \eta_{ijk}.$$

The following GLIMMIX syntax implements the fitting of the binomial data using the link function "probit."

```
proc glimmix data=germ nobound method=laplace;
class Block A B C;
model Y/N = A|B|C/link=probit;
random block block*A block*A*B;
lsmeans A|B|C/lines ilink;
run;
```

Table 6.23 shows part of the results under the binomial distribution with the "probit" link function. In parts (a) and (b), we see the mean squared error and variance component estimates for blocks, whole plot, subplot, and sub-subplot, where it can be observed that these values are positive and not negative, as the ones obtained with the link function "logit." Since the variance components are positive, this analysis makes more sense than the one based on the logit link.

The type III tests of fixed effects are tabulated in part (c) of Table 6.23; the main effects of factors A and B and the interactions *A*B*, *A*C*, and B**C* are not significant in both link functions, whereas the main effect of factor C and the interaction *A*B*C* are statistically significant under the "probit" link.

The estimated probabilities $(\hat{\pi}_{ijk})$ and their respective standard errors are presented in Table 6.24 for each of the combinations of the three factors, which

Table 6.23 Results of the analysis of variance of the RCBD in the split-split plot structure under the binomial distribution using the "probit" link

(a) Fit statistics for conditional distribution	
−2 Log L (y | r. effects)	146.43
Pearson's chi-square	43.01
Pearson's chi-square/DF $\left(\text{CME} = \hat{\sigma}^2\right)$	1.09

(b) Covariance parameter estimates		
Cov Parm	Estimate	Standard error
Block $\left(\hat{\sigma}^2_{\text{block}}\right)$	0.02411	0.03707
Block*A $\left(\hat{\sigma}^2_{\text{block}\times A}\right)$	0.02128	0.02830
Block*A*B $\left(\hat{\sigma}^2_{\text{block}\times A*B}\right)$	0.01617	0.01896

(c) Type III tests of fixed effects

				Probit	Logit
Effect	Num DF	Den DF	F-value	Pr > F	Pr > F
A	2	2	5.49	0.1541	0.4521
B	1	3	4.17	0.1339	0.9687
A*B	2	3	0.36	0.7226	0.4799
C	1	6	67.13	0.0002	0.0128
A*C	2	6	12.34	0.0075	0.3257
B*C	1	6	29.16	0.0017	0.0626
A*B*C	2	6	33.93	0.0005	0.0424

are very similar in both link functions. However, the average standard error is slightly higher with the "logit" link function $\left(\overline{\text{standar.error.mean}}_{\text{logit}} = 0.0711\right)$ compared to the "probit" link $\left(\overline{\text{standar.errormean}}_{\text{probit}} = 0.0693\right)$.

6.6.2 *Complementary Log-Log Link Function: A Split Plot in an RCBD with a Binomial Response*

Researchers studied three different micro-minerals (A, B, and C) on the attachment of explants of a commercial culture. In this vein, micro-mineral A was tested at three levels ($i = 1, 2,$ and 3), and micro-minerals B and C at two levels ($j, k =$ 1,2 and). The combination of the different levels yielded a total of 12 combinations. Since the researchers wanted to study factor C with greater precision, a split-plot treatment structure was designed in which micro-minerals A and B were placed in the whole plot (a large plot) and micro-mineral C in the subplot (a small plot). Treatment factor combinations were placed in an RCBD manner ($r = 1, 2$). The outcome of interest was the number of live plants (y_{ijkr}) out of the total number of plants growing in the

Table 6.24 Means and standard errors using the probit and logit link functions

A*B*C least squares means						
			Probit		Logit	
A	B	C	Mean	Standard error mean	Mean	Standard error mean
1	1	1	0.1543	0.05050	0.1494	0.04796
1	1	2	0.3723	0.08296	0.3780	0.08767
1	2	1	0.2724	0.06746	0.2694	0.06896
1	2	2	0.2954	0.07798	0.2953	0.08053
2	1	1	0.1023	0.03805	0.09593	0.03409
2	1	2	0.8255	0.06338	0.8292	0.06135
2	2	1	0.5684	0.08306	0.5703	0.08845
2	2	2	0.5529	0.08327	0.5530	0.08847
3	1	1	0.2844	0.07196	0.2844	0.07418
3	1	2	0.2751	0.06868	0.2733	0.07041
3	2	1	0.2568	0.06452	0.2563	0.06589
3	2	2	0.4612	0.08017	0.4608	0.08553

unit (n_{ijkr}). The data can be referred to in the Appendix (Data: Commercial crop explant attachment).

The GLMM for this experiment is described below (log-log data):

Distribution: $y_{ijkl} \mid r_l, r(a\beta)_{ijl} \sim$Binomial$(N_{ijk}, \pi_{ijk})$
$r_l \sim N\left(0, \sigma_r^2\right)$, $r(a\beta)_{ijl} \sim N\left(0, \sigma_{rab}^2\right)$,
Linear predictor: $\eta_{ijkl} = \eta + r_l + \alpha_i + \beta_j + (\alpha\beta)_{ijl} + r(\alpha\beta)_{il} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$, $_i + \beta_j + (\alpha\beta)_{ijl} + r(\alpha\beta)_{il} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$, *where blocks* (r_l) *and blocks x (A x B)* $((r(a\beta))_{ijl})$ *are assumed to contribute to the variation such that* $r_l \sim N\left(0, \sigma_r^2\right)$ *and* $r(a\beta)_{ijl} \sim N\left(0, \sigma_{rab}^2\right)$, *respectively.*
Link function: log − log $(\pi_{ijkl}) = \eta_{ijkl}$

The following GLIMMIX code adjusts the binomial proportions with a complementary link function log − log in an RCBD manner.

```
proc glimmix data=spp nobound method=laplace;
class block A B C;
model y/n = A|B|C/link=ccll;
random block block(A*B);
lsmeans A|B|C/lines ilink;
run;
```

The "link = ccll" option specifies that "proc GLIMMIX" will fit the model using the complementary (log − log) link function. The "lsmeans A|B|C/lines ilink" command calls for estimation of the linear predictors η_{ijk}, whereas the "lines" and "ilink" options provide the comparison between the linear predictors and their inverse. Part of the output is shown below. Table 6.25 shows the variance component estimates of blocks and blocks (A×B) using alternative link functions. Under

Table 6.25 Variance component estimates using the same distribution but a different link function

Covariance parameter estimates						
	Log – log		Logit		Probit	
Cov Parm	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Block	0.05808	0.07112	0.08144	0.1042	0.02676	0.03494
Block (A*B)	0.05065	0.03121	0.09203	0.05754	0.03374	0.02111

Table 6.26 Type III tests of fixed effects using the same distribution but with a different link function

Type III tests of fixed effects								
			Log – log		Logit		Probit	
Effect	Num DF	Den DF	F-value	Pr > F	F-value	Pr > F	F-value	Pr > F
A	2	5	6.27	0.0434	7.44	0.0318	8.17	0.0266
B	1	5	4.85	0.0789	3.13	0.1370	2.81	0.1543
A*B	2	5	0.65	0.5613	0.28	0.7693	0.24	0.7971
C	1	6	68.84	0.0002	65.29	0.0002	66.70	0.0002
A*C	2	6	11.94	0.0081	11.53	0.0088	12.12	0.0078
B*C	1	6	27.51	0.0019	28.88	0.0017	28.77	0.0017
A*B*C	2	6	32.44	0.0006	32.36	0.0006	33.93	0.0005

Table 6.27 Fit statistics using the same distribution but a different link function

Covariance parameter estimates			
	Log – log	Logit	Probit
−2 Log likelihood	164.85	172.57	170.88
AIC (smaller is better)	192.85	200.57	198.88
AICC (smaller is better)	239.51	247.24	245.55
BIC (smaller is better)	174.55	182.27	180.59
CAIC (smaller is better)	188.55	196.27	194.59
HQIC (smaller is better)	154.58	162.31	160.62

the link "probit," the variance components are smaller compared to those obtained with the link functions "log – log" and "logit."

The values of the hypothesis tests for the fixed effects, both main effects and interactions, are shown in Table 6.26. The three link functions behave similarly.

One tool that might be useful in choosing which link function provides a better fit, or which best describes the variability of a dataset, is the model fit statistics. The fit statistics indicate that the model with the complementary "log − log" link function provides the best fit (Table 6.27).

Table 6.28 shows the maximum likelihood estimators $\left(\hat{\pi}_{ijk}\right)$ for each of the link functions and the combination of factor levels, and it can be verified that they provide very similar estimates. It is important to mention that the correct

Table 6.28 Means and standard errors using the same distribution but with a different link function

A*B*C least squares means								
			Log − log		Logit		Probit	
A	B	C	Mean	Standard error mean	Mean	Standard error mean	Mean	Standard error mean
1	1	1	0.1494	0.04259	0.1513	0.04732	0.1547	0.05030
1	1	2	0.3776	0.08554	0.3727	0.08510	0.3696	0.08223
1	2	1	0.2661	0.06257	0.2706	0.06744	0.2737	0.06718
1	2	2	0.3001	0.07718	0.2993	0.07951	0.2980	0.07789
2	1	1	0.1020	0.03079	0.1023	0.03451	0.1047	0.03829
2	1	2	0.8389	0.08212	0.8188	0.06189	0.8196	0.06375
2	2	1	0.5558	0.09578	0.5733	0.08633	0.5700	0.08251
2	2	2	0.5578	0.09596	0.5560	0.08635	0.5546	0.08273
3	1	1	0.2770	0.06780	0.2805	0.07192	0.2827	0.07131
3	1	2	0.2782	0.06574	0.2779	0.06929	0.2778	0.06855
3	2	1	0.2555	0.05987	0.2561	0.06416	0.2569	0.06410
3	2	2	0.4599	0.08735	0.4610	0.08331	0.4609	0.07965

specification of the linear predictor as well as the distribution of the response variable are the most important elements for obtaining a good fit.

6.7 Percentages

In this section, we consider proportions that have been calculated from discrete counts, for example, the number of infected plants in treatment i of total N_i plants that are likely to have a binomial distribution. This class of models allows the response to arise from different distributions and probabilities.

6.7.1 RCBD: Dead Aphid Rate

An experiment was designed to study the effect of conidial density on the transmission of a fungus that attacks aphids. Aphid carcasses killed by the fungus, and from which the fungus released spores, were placed on bean plants at three densities ($A = 1$, $B = 5$, or $C = 10$ carcasses per plant) to provide different doses of fungal conidia. Densities were assigned to individual bean plants in a completely randomized design with six replicates. A total of 20 live uninfected (N) aphids were placed on each plant with a ladybug that was allowed to forage (feed on the bean plants) to facilitate the transfer of conidia between the carcasses and the live aphids. For each plant, the number of aphids infected with the fungus was counted (n_{ij}) and the proportion of aphids infected with the fungus was calculated 7 days after the

Table 6.29 Proportion of infested aphids

Plant	Density	p_{ij}
1	C	0.34299
2	A	0.16659
3	B	0.47004
4	C	0.62481
5	B	0.21926
6	B	0.16659
7	C	0.47502
8	C	0.52747
9	A	0.41581
10	B	0.42556
11	A	0.19466
12	A	0.34299
13	C	0.677
14	C	0.76674
15	A	0.13124
16	B	0.58419
17	B	0.38225
18	A	0.28905

Table 6.30 Sources of variation and degrees of freedom

Sources of variation	Degrees of freedom
Trt	$t - 1 = 2$
Error	$t(r - 1) = 15$
Total	$t \times r - 1 = 17$

inoculum was placed. The results shown below correspond to the proportion of infected aphids calculated at each of the inoculum concentrations ($p_{ij} = n_{ij}/N$; $N = 20$) to each of the conidial concentrations (density) tested (Table 6.29).

The sources of variation and degrees of freedom for this experiment are shown in Table 6.30.

The components of the GLMM having a beta response are listed below:

Distributions: $p_{ij} \mid \text{density(plant)}_{i(j)} \sim \text{Beta}(\pi_{ij}, \phi)$

$\text{density(plant)}_{i(j)} \sim N\left(0, \sigma^2_{\text{density(plant)}}\right)$

Linear predictor: $\eta_{ij} = \mu + \text{density}_i + \text{density(plant)}_{i(j)}$; $i = 1, 2, 3$; $j = 1, \cdots, 6$

Link function: $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \text{logit}\left(\pi_{ij}\right) = \eta_{ij}$

The following GLIMMIX program fits a GLMM in a completely randomized design with a beta distribution. Here, density is conc_ino.

```
proc glimmix data=thumbs nobound method=laplace;
class plant conc_ino;
model p = conc_ino /dist=beta link=logit;
random conc_ino(plant);
```

Table 6.31 Results of the analysis of variance

(a) Fit statistics for conditional distribution				
−2 Log L (P | r. effects)				−24.13
Pearson's chi-square				18.45
Pearson's chi-square/DF				1.02
(b) Covariance parameter estimates				
Cov Parm		Estimate		Standard error
Conc_Ino (Planta)		−0.1833		.
Scale		12.9999		4.1954
(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Conc_Ino	2	15	8.25	0.0038

Table 6.32 Means and standard errors on the model scale and the data scale

Conc_Ino least squares means							
Conc_Ino	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
A	−1.0340	0.2438	15	−4.24	0.0007	0.2623	0.04717
B	−0.5282	0.2246	15	−2.35	0.0328	0.3709	0.05241
C	0.2775	0.2197	15	1.26	0.2259	0.5689	0.05388

```
lsmeans conc_ino/lines ilink;
run;
```

Part of the results is shown in Table 6.31. The value of the conditional fit statistic in part (a), Pearson's chi − square/DF = 1.02, indicates that there is no overdispersion in the data and that the beta distribution is a good model for this dataset. The estimated variance of the plants' nested inoculum density is $\hat{\sigma}^2_{\text{density(plant)}} = -0.1833$ and the estimated scale parameter is $\hat{\phi} = 12.999$; both are tabulated in part (b). In part (c) of the same table, the type III tests of fixed effects are shown, indicating that the density (concentration) of the inoculum has a significant effect ($P = 0.0038$) on the proportion of infected aphids with the fungus.

The values under the column "Estimates" are estimated mean proportions on the model scale, whereas the column "Mean" shows the estimated mean proportions on the data scale with their respective standard errors (Table 6.32). These estimates where obtained with the "lsmeans" and "ilink" option.

Figure 6.8 shows a linear trend in the proportion of aphids infested as conidial density increases. Conidia densities A and B showed statistically equal proportions of infested aphids compared to density C. Finally, the highest proportion of infested aphids was observed at density C.

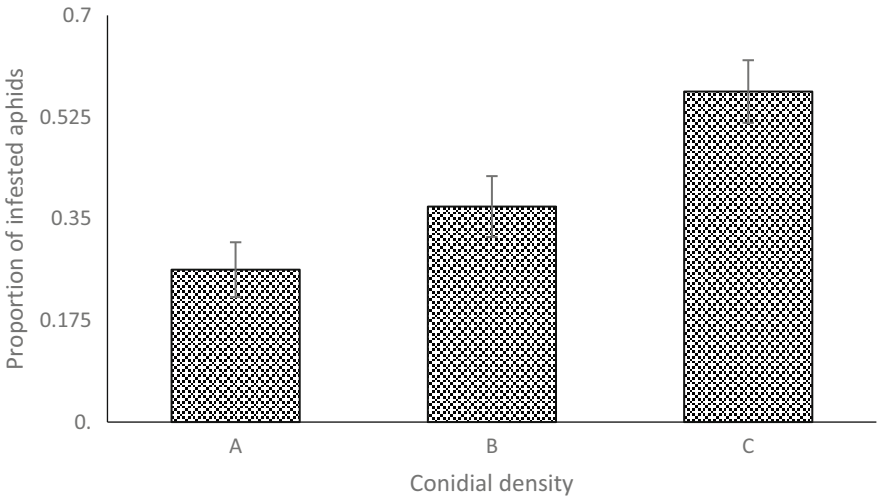

Fig. 6.8 Proportion of aphids infected at different conidia concentration densities

6.7.2 RCBD: Percentage of Quality Malt

An agro-industrial engineer is interested in studying the effect of germination time in minutes (48, 96, and 144) on the percentage of quality malt obtained from six sorghum varieties (sorghum bicolor): Gambella 1107, Macia, Meko, Red Swazi, Teshale, and 76T1#23 (Bekele et al. 2012). The percentage of quality malt (y) as a function of both factors is shown in Table 6.33.

For this purpose, an RCBD was implemented with a treatment factorial structure (variety × germination time). The statistical model to analyze the dataset is the following:

Distributions: $y_{ijk} \mid r_k \sim \text{Beta}(\pi_{ijk}, \phi)$; $i = 1, \cdots, 6$; $j, k = 1, 2, 3$
$r_k \sim N\left(0, \sigma^2_{\text{block}}\right)$, where y_{ijk} is the kth percentage of malt quality observed at the ith variety with the jth fermentation time.
Linear predictor: $\eta_{ijk} = \mu + r_k + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, where μ is the overall mean, α_i is the fixed effect due to variety i, β_j is the fixed effect due to germination time j, and $(\alpha\beta)_{ij}$ is the interaction effect between variety and germination time.
Link function: $\text{logit}(\pi_{ijk}) = \eta_{ijk}$

Table 6.34 shows the sources of variation and degrees of freedom for this experiment.

The following GLIMMIX commands adjust a GLMM with a beta response.

```
proc glimmix data=malting nobound method=laplace;
class var_sorghum ger_time block;
model p = var_sorghum|ger_time/dist=beta link=logit;
random block;
lsmeans var_sorghum|ger_time/lines ilink;
run;
```

Table 6.33 Percentage of quality malt as a function of both factors (variety and germination time)

Variety	Time	Block	y	Variety	Time	Block	y
Gambella	T1	1	7.25	Red Swazi	T2	1	21
Gambella	T1	2	11.16	Red Swazi	T2	2	15.09
Gambella	T1	3	15.9	Red Swazi	T2	3	24.84
Macia	T1	1	10.91	Teshale	T2	1	25.42
Macia	T1	2	8.75	Teshale	T2	2	26.86
Macia	T1	3	10.87	Teshale	T2	3	26.64
Meko	T1	1	24.65	76 T1#23	T2	1	23.69
Meko	T1	2	23.63	76 T1#23	T2	2	20.71
Meko	T1	3	28.75	76 T1#23	T2	3	26.14
Red Swazi	T1	1	20.95	Gambella	T3	1	12.45
Red Swazi	T1	2	15.82	Gambella	T3	2	15.34
Red Swazi	T1	3	25.24	Gambella	T3	3	17.32
Teshale	T1	1	25.92	Macia	T3	1	8.51
Teshale	T1	2	27.64	Macia	T3	2	8.15
Teshale	T1	3	28.03	Macia	T3	3	13.07
76T1#23	T1	1	23.39	Meko	T3	1	22.09
76T1#23	T1	2	19.43	Meko	T3	2	24.11
76T1#23	T1	3	25.55	Meko	T3	3	24.47
Gambella	T2	1	10.03	Red Swazi	T3	1	20.81
Gambella	T2	2	12.9	Red Swazi	T3	2	16.05
Gambella	T2	3	17.84	Red Swazi	T3	3	23.7
Macia	T2	1	7.88	Teshale	T3	1	26.42
Macia	T2	2	9.14	Teshale	T3	2	27.07
Macia	T2	3	11.99	Teshale	T3	3	28.01
Meko	T2	1	22.97	76 T1#23	T3	1	24.18
Meko	T2	2	25.37	76 T1#23	T3	2	19.58
Meko	T2	3	25.71	76 T1#23	T3	3	25.74

Table 6.34 Sources of variation and degrees of freedom

Sources of variation	Degrees of freedom
Blocks	$r - 1 = 3 - 1 = 2$
Variety	$a - 1 = 6 - 1 = 5$
Time_Germination	$b - 1 = 3 - 1 = 2$
Variety*germ_time	$(a - 1)(b - 1) = 10$
Error	$(ab - 1)(r - 1) = 17 \times 2 = 34$
Total	$r \times a \times b - 1 = 54 - 1 = 53$

Part of the results of the above program is shown in Table 6.35. In part (a), the value of Pearson's chi-square/DF is tabulated $\left(\frac{\chi^2}{df} = 0.92\right)$, which indicates that the beta distribution is a good distribution for modeling malt percentage since the t-value of Pearson's chi-square/DF is close to 1. The estimated variance due to blocks is

Table 6.35 Results of the analysis of variance of the RCBD with a beta distribution

(a) Fit statistics for conditional distribution				
−2 Log L (p | r. effects)				−280.89
Pearson's chi-square				49.66
Pearson's chi-square/DF				0.92
(b) Covariance parameter estimates				
Cov Parm	Estimate			Standard error
Block	0.01210			0.01055
Scale	431.54			85.4922
(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Var_sorghum	5	34	106.51	<0.0001
Ger_time	2	34	0.26	0.7722
Var_sorghum*ger_time	10	34	1.08	0.4041

Table 6.36 Means and standard errors on the model scale and the data scale for sorghum varieties

Var_sorghum least squares means							
Var_sorghum	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
76 T1#23	−1.2011	0.07401	34	−16.23	<0.0001	0.2313	0.01316
Gambella	−1.8898	0.07929	34	−23.83	<0.0001	0.1313	0.009042
Macia	−2.2067	0.08295	34	−26.60	<0.0001	0.09915	0.007409
Meko	−1.1201	0.07364	34	−15.21	<0.0001	0.2460	0.01366
Red Swazi	−1.3685	0.07493	34	−18.26	<0.0001	0.2029	0.01212
Teshale	−1.0025	0.07314	34	−13.71	<0.0001	0.2685	0.01436

$\hat{\sigma}^2_{\text{block}} = 0.012$ and the estimated scale parameter is $\hat{\phi} = 431$ (part (b)), whereas the type III fixed effects hypothesis tests in part (c) show that sorghum variety has a significant effect on malt quality percentage ($P = 0.0001$).

The least squares means on the model scale and the data scale for the factor variety are listed under the columns "Estimate" and "Mean" with their respective standard errors "Standard error" in Table 6.36.

Figure 6.9 shows that Teshale produced the highest average malt percentage (0.2685 ± 0.01436), followed by the varieties 76 T1#23 and Meco (0.2313 ± 0.01316,0.246 ± 0.01366), whereas the variety Macia produced the lowest malt percentage (0.09915 ± 0.0074).

*6.7.3 A Split Plot in an RCBD: Cockroach Mortality (***Blattella germanica***)*

An entomologist is interested in testing six isolates of insect pathogenic fungi: five obtained from different hosts and one already known isolate (Control) of a fungus

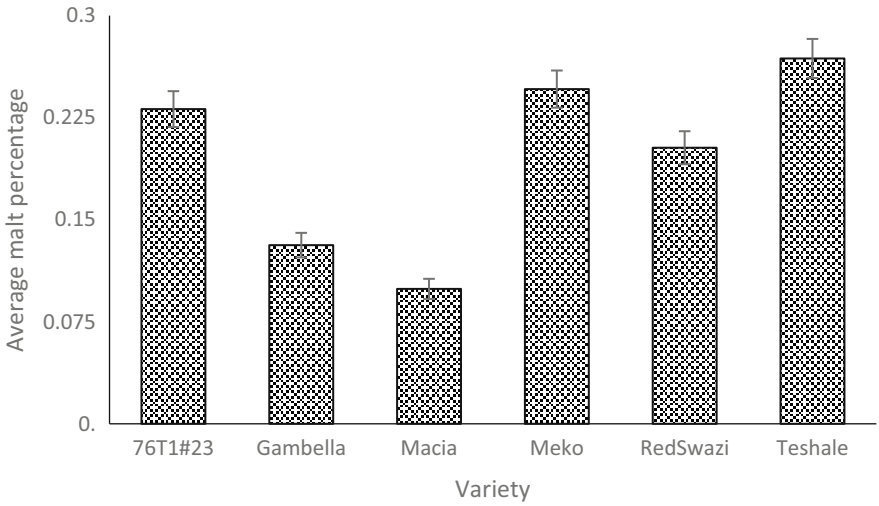

Fig. 6.9 Percentage of quality malt of bicolor sorghum varieties

Table 6.37 Analysis of variance with sources of variation and degrees of freedom for this experiment

Sources of variation	Degrees of freedom
Blocks	$r - s1 = 2 - 1 = 1$
Isolation	$a - 1 = 6 - 1 = 5$
Block (insulation)	$a(r - 1) = 6$
Age	$b - 1 = 3 - 1 = 2$
Isolation*age	$(a - 1)(b - 1) = 5 \times 2 = 10$
Error	$(a - 1)(b - 1)(r - 1) = 2 \times 5 \times 1 = 10$
Total	$r \times a \times b - 1 = 2 \times 6 \times 3 - 1 = 35$

with potential for biological control of a particular species of cockroaches. To do so, the entomologist decides to test these fungal isolates on three different insect ages (age1 = E1, age2 = E2, and age3 = E3). Each of the isolates was placed in a Petri dish with 10 insects of a specific age. Each set (isolate–age) was randomly assigned to two blocks (Appendix: Data: Cockroaches).

The analysis of variance table (Table 6.37) with the sources of variation and degrees of freedom for this experiment is presented below. The response variable (percentage mortality) for this experiment is assumed to have a beta distribution.

The components that describe the model of this experiment are listed below:

Distributions: $y_{ijk} \mid r_k, r(\alpha)_{k(i)} \sim \text{Beta}(\pi_{ijk}, \phi)$; $i = 1, \cdots, 6$; $j = 1, 2, 3$; $k = 1, 2$.

$r_k \sim N\left(0, \sigma_r^2\right), r(\alpha)_{k(i)} \sim N\left(0, \sigma_{r(\alpha)}^2\right)$

Linear predictor: $\eta_{ijk} = \mu + r_k + \alpha_i + r(\alpha)_{k(i)} + \beta_j + (\alpha\beta)_{ij}$

Link function: $logit(\pi_{ijk}) = \eta_{ijk}$

Table 6.38 Results of the analysis of variance of the RCBD with a factorial structure in treatments

(a) Fit statistics for conditional distribution			
−2 Log L (y | r. effects)			−74.53
Pearson's chi-square			34.02
Pearson's chi-square/DF			1.00
(b) Covariance parameter estimates			
Cov Parm	Subject	Estimate	Standard error
Aislamiento	Block	−0.03125	.
Scale		24.1882	5.7925
(c) Type III tests of fixed effects			
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Isolation	5	6	16.48	0.0019
Age	2	10	30.01	<0.0001
Isolation*age	10	10	4.83	0.0102

The following GLIMMIX commands adjust a GLMM with a beta response.

```
proc glimmix nobound method=laplace;
class block Isolation Age;
model y = Isolation|Age/dist=beta link=logit;
random Isolation/subject=block;
lsmeans Insulation|Age/slice=Insulation lines ilink;
run;
```

Some of the outputs are listed below (Table 6.38). The conditional statistic Pearson's chi − square/DF = 1 indicates that the distribution used is appropriate for these datasets (part (a)). The variance component estimates are tabulated in part (b), and, for blocks, the estimate is $\hat{\sigma}_r^2 = -0.03125$ and the estimated scale parameter is $\hat{\phi} = 24.1882$. The hypothesis test is in part (c) with type III fixed effects of equality of means for type of isolation, age of the insect, and the interaction between both factors. These outputs indicate that they have a significant effect on insect mortality.

We see the expected proportions with their respective standard errors of both factors on the data scale under the "Mean" column (Tables 6.39 and 6.40). These values arise by applying the inverse link to estimates under "Estimate" on the model scale. Table 6.39 shows the estimated average mortality probabilities for the isolates; for example, for isolate A1, applying the inverse link to the linear predictor estimate $\hat{\eta}_{1.} = 0.1722$ we get $\hat{\pi}_{1.} = 1/1 + e^{-0.1722} = 0.5429$. In this manner, we see that the expected proportions for isolates 2 and 4 are $\hat{\pi}_{2.} = 0.6555$ and $\hat{\pi}_{4.} = 0.5762$, respectively, whereas for the control $\hat{\pi}_{\text{control.}} = 0.1157$.

Regarding the age of the insect (Table 6.40), the expected average probability of mortality was higher at age three (adults) with a higher mortality rate $\hat{\pi}_{.3} = 0.6435$, whereas insects at age two (E2) had a higher resistance to the isolations, showing a mortality of $\hat{\pi}_{.2} = 0.2598$.

In general, fungal isolates A1, A2, A3, and A4 showed an average mortality of more than 75% for adult insects (E3), whereas isolates A1, A2, and A5 showed a

Table 6.39 Means and standard errors on the model scale and the data scale for isolation

Isolate least squares means							
Isolate	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
A1	0.1722	0.1859	6	0.93	0.3900	0.5429	0.04614
A2	0.6442	0.2100	6	3.07	0.0220	0.6557	0.04740
A3	−0.1489	0.1952	6	−0.76	0.4746	0.4629	0.04853
A4	0.3073	0.2088	6	1.47	0.1915	0.5762	0.05098
A5	−0.2023	0.1806	6	−1.12	0.3053	0.4496	0.04468
Control	−2.0339	0.2418	6	−8.41	0.0002	0.1157	0.02473

Table 6.40 Means and standard errors on the model scale and the data scale for insect age

Age least squares means							
Age	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
E1	−0.1747	0.1310		−1.33	0.2120	0.4564	0.03251
E2	−1.0468	0.1374		−7.62	<0.0001	0.2598	0.02643
E3	0.5908	0.1634		3.61	0.0047	0.6435	0.03749

mortality rate of around 65% for cockroaches of age E1 (juvenile insects). On the other hand, all isolates showed lower lethal effectiveness on insects of age E2 (Fig. 6.10).

6.7.4 *A Split-Plot Design in an RCBD: Percentage Disease Inhibition*

A plant pathologist wishes to compare the response of two plant varieties to different doses/amounts of a pesticide formulated to protect plants against a disease. Five racks (blocks) were chosen to account for local variation within the greenhouse. Each rack was divided into four sections or rooms and were randomly assigned one of four pesticide levels to each rack. The four pesticide levels were 1, 2, 4, and 8 mg/L. One plant of each variety was placed in each section of the rack. Of the two plant varieties, one variety was susceptible, labeled S, and the other variety was resistant, labeled R (Table 6.41). The response variable (y) is the percentage of disease inhibition in the plant.

The sources of variation and degrees of freedom for this experiment are shown in Table 6.42.

Following the same reasoning used in the examples above, the components of the GLMM with a beta response that models the observed disease inhibition proportion (p_{ijk}) under dose i with variety j in block k are listed as follows:.

Distributions: $y_{ijk} \mid r_k, (\mathrm{r}\alpha)_{ik} \sim \mathrm{Beta}(\pi_{ijk}, \phi)$; $i = 1, \cdots, 4$; $j = 1, 2$; $k = 1, \cdots, 5$

$r_k \sim N\left(0, \sigma_{\mathrm{r}}^2\right)$, $(r\alpha)_{ik} \sim N\left(0, \sigma_{\mathrm{rA}}^2\right)$

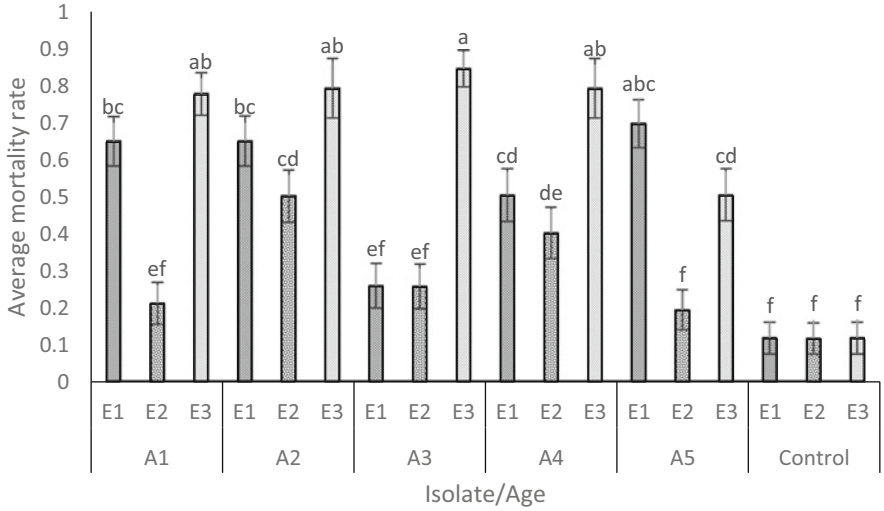

Fig. 6.10 Cockroach mortality percentage

Table 6.41 Percentage of inhibition

Block	Variety	Dose	*y*	Block	Variety	Dose	*y*
1	R	1	15.7	1	S	1	19.8
2	R	1	23.1	2	S	1	17.8
3	R	1	15.9	3	S	1	13.2
4	R	1	20.8	4	S	1	14.8
5	R	1	24.5	5	S	1	19.7
1	R	2	25.1	1	S	2	21.2
2	R	2	29.2	2	S	2	29.3
3	R	2	29.7	3	S	2	26
4	R	2	28.6	4	S	2	27.5
5	R	2	26.6	5	S	2	22
1	R	4	27.9	1	S	4	29.3
2	R	4	29.7	2	S	4	27.2
3	R	4	24	3	S	4	26
4	R	4	29.7	4	S	4	31.5
5	R	4	29.6	5	S	4	27.9
1	R	8	23.8	1	S	8	22.8
2	R	8	31.2	2	S	8	33
3	R	8	21.8	3	S	8	25.2
4	R	8	23.3	4	S	8	27.2
5	R	8	23.9	5	S	8	20.8

Table 6.42 Sources of variation and degrees of freedom

Sources of variation	Degrees of freedom
Blocks	$r - 1 = 5 - 1 = 4$
Dose	$a - 1 = 4 - 1 = 3$
Error_a(Bloque*Dose)	$(r - 1)(a - 1) = 12$
Variety	$b - 1 = 2 - 1 = 1$
Dose*variety	$(a - 1)(b - 1) = 3$
Error_b	$a(b - 1)(r - 1) = 4 \times 1 \times 4 = 16$
Total	$r \times a \times b - 1 = 5 \times 4 \times 2 - 1 = 39$

Linear predictor: $\eta_{ijk} = \mu + r_k + \alpha_i + (r\alpha)_{ik} + \beta_j + (\alpha\beta)_{ij}$, where r_k is the random block effect, α_i is the fixed dose effect, β_j is the fixed variety effect, $(r\alpha)_{ik}$ is the random effect due to block by dose interaction, and $(\alpha\beta)_{ij}$ is the interaction of fixed effects due to dose variety.

Link function: $\text{logit}(\pi_{ijk}) = \eta_{ijk}$

The following GLIMMIX commands adjust a GLMM.

```
proc glimmix nobound method=laplace;
class Variety dose block;
model y = dose variety dose*variety /dist=beta link=logit;
random Block Block*dose;
contrast 'Linear dose' dose -3 -1 1 3;
contrast 'Quadratic dose' dose 1 -1 -1 -1 1;
contrast 'dose Cubic' dose -1 3 -3 1;
lsmeans variety|dose / slice=(variety dose) lines ilink;
ods output lsmeans=dose_means;
run;
```

The "contrast" command in the program can perform a hypothesis testing to see what trend (linear, quadratic, or cubic) the "dose" factor has on the percentage of disease inhibition. Part of the output is shown in Table 6.43. The value of the conditional goodness-of-fit statistic Pearson's chi − square/DF= 0.59 indicates that we have no evidence of overdispersion, and, therefore, the beta distribution is adequate to model this dataset (part (a)). The variance component estimates in part (b) for block and block × dose are $\hat{\sigma}^2_{\text{r}} = 0.004898$ and $\hat{\sigma}^2_{\text{r}\cdot\text{dose}} = 0.002372$, respectively. Finally, the F-value provides sufficient statistical evidence of the effect of dose on disease decline in plants ($P = 0.0001$), whereas the effect of variety and dose × variety do not provide sufficient evidence.

Table 6.44 shows the polynomial contrasts for the effect of "dose," which indicate that there is a significant quadratic effect on the percentage of disease inhibition.

The inhibition percentage has almost a linear trend as the dose increases from 1 to 4 ml/L in both varieties, but when the dose is higher than 4 ml/L, the inhibition of the disease decreases in both varieties (Fig. 6.11).

Table 6.43 Results of the analysis of variance

(a) Fit statistics for conditional distribution				
−2 Log L (y | r. effects)			−184.32	
Pearson's chi-square			23.63	
Pearson's chi-square/DF			0.59	
(b) Covariance parameter estimates				
Cov Parm	Estimate		Standard error	
Block	0.004898		.	
Block*dose	0.002372		0.007513	
Scale	205.52		67.7447	
(c) Type III tests of fixed effects				
Effect	Num DF	Den DF	*F*-value	Pr > *F*
Dose	3	12	17.67	0.0001
Variety	1	16	1.74	0.2057
Dose*variety	3	16	1.22	0.3337

Table 6.44 Polynomial contrasts

Contrasts				
Label	Num DF	Den DF	*F*-value	Pr > *F*
Linear dose	1	12	25.48	0.0003
Quadratic dose	1	12	30.93	0.0001
Cubic dose	1	12	0.30	0.5948

6.7.5 *Randomized Complete Block Design with a Binomial Response with Multiple Variance Components*

The dataset corresponds to an experiment implemented by Madden and Hughes (1995) on the incidence of the disease caused by the fungus *Plasmopara viticola* on grape plants (*Vitis labrusca*). Six different treatments in a randomized block design ($b = 3$) were tested, where treatment 1 was the control, to study the disease with three grape plants ($v = 3$). On a single date in autumn, five sprouts were ($r = 5$) randomly selected from each of the three grape plants and the number of leaves with at least one mildew lesion was counted (m) out of a total n leaves. The number of leaves per shoot ranged from 7 to 21. The data for this experiment can be found in the Appendix (Data: Disease incidence on grape plants).

The statistical model that could describe the incidence of disease in this experiment, if the response variable p_{ijkl} were treated as a normal variable, would be as described below:

$$p_{ijkl} = \eta + \tau_i + b_j + (bv)_{jk} + (bvr)_{jkl} + \varepsilon_{ijkl}$$

$$i = 1, 2, \ldots, 6; j = 1, 2, 3; k = 1, 2, .., 3; l = 1, 2, \ldots, 5$$

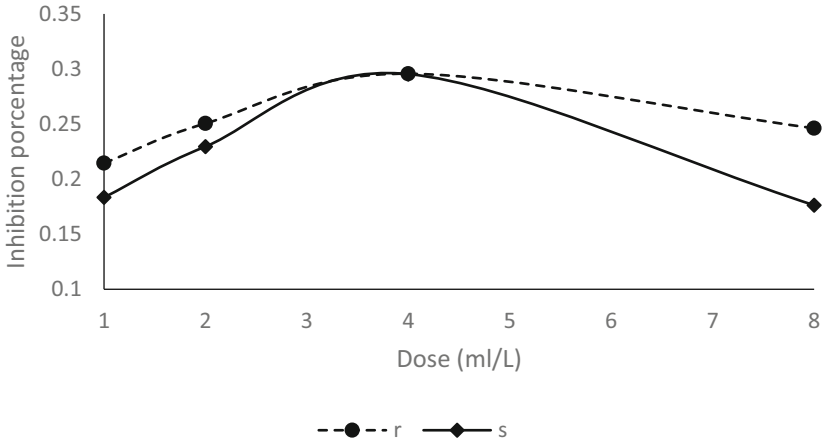

Fig. 6.11 Percentage of disease inhibition in both varieties

where p_{ijkl} is the *ijkl* proportion of diseased leaves, η is the intercept, τ_i is the fixed treatment effect *i*, b_j is the random effect of blocks assuming $b_j \sim N\left(0, \sigma^2_{\text{block}}\right)$, $(\text{bv})_{jk}$ is the block–plant random effect assuming $(\text{bv})_{jk} \sim N\left(0, \sigma^2_{\text{block} \times \text{plant}}\right)$, $(\text{bvr})_{jkl}$ is the random effect due to block–plant–sprouts assuming $(\text{bvr})_{jkl} \sim N\left(0, \sigma^2_{\text{block} \times \text{plant} \times \text{sprout}}\right)$, and ε_{ijkl} is the experimental error assuming $\varepsilon_{ijkl} \sim N(0, \sigma^2)$.

For the disease incidence data, the assumption of a normal distribution for p_{ijkl} is not recommended. A good starting point for the analysis is to assume that the observed number of diseased leaves in the sprouts (y_{ijkl}) follows a binomial distribution with parameter π_{ijkl} and n_{ijkl}, the total number of leaves on the sprout.

Therefore, the components of the GLMM with a binomial distribution in the response variable are as follows:

Distribution: $p_{ijkl} \mid b_j, (\text{bv})_{jk}, (\text{bvr})_{jkl} \sim \text{binomial}(\pi_{ijkl}, n_{ijkl})$
$b_j \sim N\left(0, \sigma^2_{\text{block}}\right), (\text{bv})_{jk} \sim N\left(0, \sigma^2_{\text{block} \times \text{plant}}\right), (\text{bvr})_{jkl} \sim N\left(0, \sigma^2_{\text{block} \times \text{plant} \times \text{sprout}}\right)$
Linear predictor: $\eta_{ijkl} = \eta + \tau_i + b_j + (\text{bv})_{jk} + (\text{bvr})_{jkl.}$
Link function: $\text{logit}(\pi_{ijkl}) = \eta_{ijkl}$

The following GLIMMIX syntax fits a GLMM with a binomial response.

```
proc glimmix method=laplace nobound;
class v r b t;
model m/n = t /dist=bin;
random intercept v v*r/subject=b;
lsmeans t/lines ilink;
run;
```

Table 6.45 Results of the analysis of variance under the binomial distribution

(a) Fit statistics	
−2 Log likelihood	723.17
AIC (smaller is better)	741.17
AICC (smaller is better)	741.87
BIC (smaller is better)	733.06
CAIC (smaller is better)	742.06
HQIC (smaller is better)	724.87

(b) Fit statistics for conditional distribution	
−2 Log L (m | r. effects)	665.02
Pearson's chi-square	398.21
Pearson's chi-square/DF	1.47

(c) Covariance parameter estimates			
Cov Parm	Subject	Estimate	Standard error
Intercept	b	−0.00408	.
V	b	0.01917	.
v*r	b	0.1960	.

(d) Type III tests of fixed effects				
Effect	Num DF	Den DF	F-value	Pr > F
t	2	220	1837.99	<0.0001

Part of the results based on the aforementioned model is shown in Table 6.45. By default, proc GLIMMIX provides the fit statistics useful for selecting the best model from a group of models (part (a)).

In addition to accuracy considerations, the Laplace (or quadrature) analysis allows us to obtain the "conditional distribution fit statistics," specifically Pearson's χ^2/df. Recall that this statistic helps assess the goodness of fit of the model. If the value of $\chi^2/df \gg 1$ is an indicator that there is overdispersion in the dataset, then this may be because the linear predictor is incomplete or the assumed distribution is not suitable (mis-specified) for this dataset. In part (b), we can see that the value of the conditional distribution statistic of Pearson's $\chi^2/df = 1.47$. This value indicates that we have evidence of overdispersion. The variance component estimates due to block, block × plant, and block × plant × sprout are tabulated in part (c), whereas the type III tests of fixed effects (part (d)) indicate that there is a significant difference ($P < 0.0001$) between treatments.

Since there is overdispersion in the data in the binomial model, an alternative distribution is the beta distribution. The components of the GLMM are as follows:

Distribution: $p_{ijkl} \mid b_j, (\text{bv})_{jk}, (\text{bvr})_{jkl} \sim \text{beta}(\pi_{ijkl}, \phi)$;

$b_j \sim N\left(0, \sigma^2_{\text{block}}\right), (\text{bv})_{jk} \sim N\left(0, \sigma^2_{\text{block} \times \text{plant}}\right), (\text{bvr})_{jkl} \sim N\left(0, \sigma^2_{\text{block} \times \text{plant} \times \text{sprout}}\right)$

Linear predictor: $\eta_{ijkl} = \eta + \tau_i + b_j + (\text{bv})_{jk} + (\text{bvr})_{jkl}$

Link function: $\text{logit}(\pi_{ijkl}) = \eta_{ijkl}$.

Table 6.46 Results of the analysis of variance under the beta distribution

(a) Fit statistics	
−2 Log likelihood	−231.10
AIC (smaller is better)	−211.10
AICC (smaller is better)	−209.30
BIC (smaller is better)	−220.11
CAIC (smaller is better)	−210.11
HQIC (smaller is better)	−229.22

(b) Fit statistics for conditional distribution	
−2 Log L (m | r. effects)	−231.10
Pearson's chi-square	136.55
Pearson's chi-square/DF	1.07

(c) Covariance parameter estimates

Cov Parm	Subject	Estimate	Standard error
Intercept	b	−0.6308	.
V	b	−0.2215	.
v*r	b	−0.1843	.
Scale $(\hat{\phi})$		9.8397	1.1926

(d) Type III tests of fixed effects

Effect	Num DF	Den DF	F-value	Pr > F
t	2	220	1837.99	<0.0001

The following SAS commands adjust an GLMM under a beta distribution.

```
proc GLIMMIX method=laplace nobound;
class v r b t;
model pct = t /dist=beta link=logit;
random intercept v v*r/subject=b;
lsmeans t/lines ilink;
run;
```

Some of the outputs are shown below. Table 6.46 shows that the values of the fit statistics, as well as the conditional distribution statistics (parts (a) and (b)), are much smaller than when the binomial distribution was used.

This indicates that the beta distribution is more appropriate for the dataset, as the value of Pearson's statistic is $\chi^2/df = 1.03$, indicating that the problem of overdispersion was almost totally controlled. The variance component estimates as well as the estimated scale parameter $(\hat{\phi})$ are tabulated in part (c). Similar to the previous analysis, the type III tests of fixed effects indicate that there is a highly significant difference (part (d)) in treatments on the average proportion of leaves with fungal disease.

The least mean squares (means) on the model scale (column "Estimate") and on the data scale (column "Mean") are tabulated in Table 6.47. The results indicate that

Table 6.47 Estimated means (least squares means) on the model scale and on the data scale

Least squares means							
t	Estimate	Standard error	DF	*t*-value	Pr > |*t*|	Mean	Standard error mean
1	0.7223	0.09989	83	7.23	<0.0001	0.6731	0.02198
2	−1.7482	0.1543	83	−11.33	<0.0001	0.1483	0.01949
3	−2.0178	0.2214	83	−9.11	<0.0001	0.1174	0.02294
4	−1.9358	0.1873	83	−10.34	<0.0001	0.1261	0.02064
5	−1.7887	0.2173	83	−8.23	<0.0001	0.1432	0.02667
6	−1.5360	0.1665	83	−9.23	<0.0001	0.1771	0.02427

Table 6.48 Mean comparison (LSD method)

T grouping of *t* least squares means ($\alpha = 0.05$)		
LS means with the same letter are not significantly different		
t	Estimate	
1	0.7223	A
6	−1.5360	B
		B
2	−1.7482	B
		B
5	−1.7887	B
		B
4	−1.9358	B
		B
3	−2.0178	B

all proposed treatments in this study reduce the proportion of diseased leaves compared to the control treatment ($t = 1$).

The mean comparison (LSD) obtained with the option "lines" indicates that the proportion of diseased leaves in treatment one is statistically different from the rest of the treatments (Table 6.48).

6.8 Exercises

Exercise 6.8.1 Seeds of a particular crop were stored at four different temperatures (T_1, T_2, T_3, and T_4) under four different chemical concentrations (0, 0.1, 1.0, and 10). To study the effects of temperature and chemical concentration, a completely randomized experiment was conducted with a factorial treatment structure 4 × 4 and four replicates. For each of the 64 experimental units, 50 seeds were placed in a dish and the number of seeds that germinated under standard conditions was recorded. Germination data were obtained from Mead et al. (1993, p. 325) (Table 6.49).

Table 6.49 Seed germination experiment results

	Chemical concentration			
Temperature	0	0.1	1.0	
T_1	9, 9, 3, 7	13, 12, 14, 15	21, 23, 24, 27	40, 32, 43, 34
T_2	19, 30, 21, 29	33, 32, 30, 26	43, 40, 37, 41	48, 48, 49, 48
T_3	7, 7, 2, 5	1, 2, 4, 4	8, 10, 6, 7	3, 4, 8, 5
T_4	4, 9, 3, 7	13, 6, 15, 7	16, 13, 18, 19	13, 18, 11, 16

Table 6.50 Results of the apple sprouts experiment

Density of inoculum	Cultivate	Block 1	Block 2	Block 3	Block 4
200	Jonagold	5/1	5/2	5/1	5/0
200	Golden delicious	5/1	5/0	5/0	5/0
200	Jonathan	5/2	5/2	5/2	5/0
1000	Jonagold	5/0	5/2	5/2	5/4
1000	Golden delicious	5/0	5/0	5/2	5/0
1000	Jonathan	5/4	5/4	5/4	5/0
5000	Jonagold	5/5	5/5	5/4	5/5
5000	Golden delicious	5/5	5/4	5/3	5/5
5000	Jonathan	5/5	5/0	5/3	5/5

The first number refers to the number of inoculations (n) and the second to the number of inoculations that developed the gangrenous sore (Y)

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for this experiment.
(b) List all the components of the GLMM in (a).
(c) Analyze this dataset and summarize the relevant results.

Exercise 6.8.2 Data were obtained from an experiment in which separate sprouts of apple trees were inoculated with macroconidia of the fungus *Nectria galligena*, which causes apple cancer (canker gangrene). The experimental factors were inoculum density (three levels: 200, 1000, and 5000 macroconidia per ml) and variety (three levels: Jonagold, Golden Delicious, and Jonathan). The experiment was carried out in 4 randomized blocks with 12 plots. Each plot consisted of one sprout on which five inoculations were made. The numbers of successful inoculations per plot on day 17 after inoculation are shown in the table below (Table 6.50).

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for this experiment.
(b) List all the components of the GLMM from part (a).
(c) Analyze this dataset and summarize the relevant results.
(d) Is there is an extra-variation in the dataset? What alternative distribution do you propose? Reanalyze the data and compare the results.

Exercise 6.8.3 This experiment concerns the germination efficiencies of protoplasts obtained from plants of seven species of the genera *Lycopersicon* (tomato) and

Table 6.51 Protoplast germination experiment results

Species	Isolation	1	2	3	4	5	6	7	8	9	10
1	1	8.9	6.3	10.5							
1	2	3.1	2.7	4.1							
1	3	2.1	1.9	1.4	1.5						
1	4	2.5	2.9	2.6	2.6	2.6	2.6	2.8	2.7	2.8	2.7
2	1	0.2	0.9	0.5	0.6	1.2	0.4				
2	2	1.8	1.6	1.6							
2	3	6.6	7.5	5.4	5.3	5	6.5	6.3	5.8	5.9	5.6
3	1	1.8	1.5	1.9	1.7	1.3	1.5				
3	2	1.5	3.2	1.1	1.3	1.8	1.2	1.6	1.4	1.2	1.8
3	3	2	2.3	2.8	2.6	3.2	2.2	2.5	2.4	2.8	2.4
4	1	11.4	11.3	14.4	13.7						
4	2	2.9	3.8	4.7	5.1	2.7	3.2				
4	3	2.3	4.4	4.8	4.9	5.8	4.7	5.6	4.2	3.3	4.5
5	1	21.5	25.5	18.1	22.2						
5	2	18.7	20								
5	3	11.5	13.1	11.5	16.2	10.1	17.2	16	10.5		
6	1	4.6	3.4	2.7	3	4.1	3.1				
6	2	2.4	2.4	2	2.5	3.6	3.2	2.6	1.4	2.5	2.7
6	3	1.6	1.1	1.6	1.3	1.6	1	0.8	1.3	0.8	2.2
7	1	3	4	4.1	4.4	2.8	3.3	4.5	3.3	3	3.2
7	2	2.5	2.5	2.5	2.7	2.3	2.6				
7	3	2.6	2.7	2.9	2.7	2.7	2.6				
7	4	2.9	3	3	3.1						

Solanum (potato). For each species, three or four protoplast isolates were used and, depending on the availability of the protoplasts, a variable number of plates was carried out. Per plate, approximately 105 protoplasts were placed in a Petri dish, and, after 4 weeks, the proportion of dividing protoplasts was recorded. The results in percentages are listed below (Table 6.51).

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for the experimental design of this study.
(b) Write down a generalized linear mixed model base in (a), assuming a beta distribution on the response variable.
(c) Implement an analysis of these data according to the linear predictor and model in part (b). Summarize the relevant results.

Exercise 6.8.4 The data in this example are the results of a triangle test for 12 raters tasting 10 pairs of coffee varieties (Table 6.52). The triangle test consisted of each rater drinking three cups, one of one variety and two of the other. Each rater had 12 triangles for each pair of varieties, 2 for each of the following sequences: AAB, ABA, BAA, ABB, BAB, and BBA. The answer is the correct variety identification number appearing once. The experiment was conducted in two groups of six

Table 6.52 Triangle test (G = group, Eval = panelist, PdV = variety pair, V_A = variety A; V_B = variety B; Y = number of correct discriminations, n = number of trials)

G	Eval	PdV	V_A	V_B	Y	n	G	Eval	PdV	V_A	V_B	Y	n
1	1	1	8	9	2	12	2	7	1	8	9	4	12
1	1	2	5	9	11	12	2	7	2	5	9	12	12
1	1	3	9	6	9	12	2	7	3	9	6	7	12
1	1	4	6	5	6	12	2	7	4	6	5	9	12
1	1	5	6	8	8	12	2	7	5	6	8	10	12
1	1	6	5	8	9	12	2	7	6	5	8	5	12
1	1	7	7	8	6	12	2	7	7	7	8	9	12
1	1	8	7	9	8	12	2	7	8	7	9	9	12
1	1	9	7	5	11	12	2	7	9	7	5	7	12
1	1	10	7	6	5	12	2	7	10	7	6	5	12
1	2	1	8	9	5	12	2	8	1	8	9	2	12
1	2	2	5	9	8	12	2	8	2	5	9	10	12
1	2	3	9	6	8	12	2	8	3	9	6	8	12
1	2	4	6	5	9	12	2	8	4	6	5	9	12
1	2	5	6	8	10	12	2	8	5	6	8	8	12
1	2	6	5	8	11	12	2	8	6	5	8	9	12
1	2	7	7	8	8	12	2	8	7	7	8	4	12
1	2	8	7	9	9	12	2	8	8	7	9	6	12
1	2	9	7	5	8	12	2	8	9	7	5	10	12
1	2	10	7	6	7	12	2	8	10	7	6	7	12
1	3	1	8	9	4	12	2	9	1	8	9	3	12
1	3	2	5	9	9	12	2	9	2	5	9	11	12
1	3	3	9	6	9	12	2	9	3	9	6	11	12
1	3	4	6	5	11	12	2	9	4	6	5	8	12
1	3	5	6	8	8	12	2	9	5	6	8	8	12
1	3	6	5	8	10	12	2	9	6	5	8	11	12
1	3	7	7	8	3	12	2	9	7	7	8	5	12
1	3	8	7	9	7	12	2	9	8	7	9	4	12
1	3	9	7	5	10	12	2	9	9	7	5	11	12
1	3	10	7	6	9	12	2	9	10	7	6	8	12
1	4	1	8	9	7	12	2	10	1	8	9	7	12
1	4	2	5	9	10	12	2	10	2	5	9	9	12
1	4	3	9	6	7	12	2	10	3	9	6	5	12
1	4	4	6	5	8	12	2	10	4	6	5	11	12
1	4	5	6	8	7	12	2	10	5	6	8	5	12
1	4	6	5	8	8	12	2	10	6	5	8	10	12
1	4	7	7	8	7	12	2	10	7	7	8	7	12
1	4	8	7	9	6	12	2	10	8	7	9	8	12
1	4	9	7	5	10	12	2	10	9	7	5	6	12
1	4	10	7	6	7	12	2	10	10	7	6	9	12
1	5	1	8	9	6	12	2	11	1	8	9	7	12
1	5	2	5	9	10	12	2	11	2	5	9	9	12

(continued)

Table 6.52 (continued)

G	Eval	PdV	V_A	V_B	*Y*	*n*	G	Eval	PdV	V_A	V_B	*Y*	*n*
1	5	3	9	6	4	12	2	11	3	9	6	6	12
1	5	4	6	5	8	12	2	11	4	6	5	10	12
1	5	5	6	8	6	12	2	11	5	6	8	5	12
1	5	6	5	8	7	12	2	11	6	5	8	10	12
1	5	7	7	8	8	12	2	11	7	7	8	8	12
1	5	8	7	9	9	12	2	11	8	7	9	6	12
1	5	9	7	5	9	12	2	11	9	7	5	9	12
1	5	10	7	6	8	12	2	11	10	7	6	9	12
1	6	1	8	9	3	12	2	12	1	8	9	6	12
1	6	2	5	9	9	12	2	12	2	5	9	7	12
1	6	3	9	6	6	12	2	12	3	9	6	7	12
1	6	4	6	5	9	12	2	12	4	6	5	7	12
1	6	5	6	8	7	12	2	12	5	6	8	8	12
1	6	6	5	8	10	12	2	12	6	5	8	11	12
1	6	7	7	8	7	12	2	12	7	7	8	9	12
1	6	8	7	9	7	12	2	12	8	7	9	9	12
1	6	9	7	5	8	12	2	12	9	7	5	10	12
1	6	10	7	6	9	12	2	12	10	7	6	9	12

evaluators, each with the aim of discriminating the abilities of the panelists for future evaluations. The data for this example are shown below:

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for this experiment.
(b) List all the components of the GLMM according to part (a).
(c) Analyze this dataset and summarize the relevant results.
(d) Is there an extra-variation in the dataset? If so, what alternative distribution do you propose? Reanalyze the data and compare the results.

Exercise 6.8.5 Several brewing techniques are used in the production of espresso coffee. Among them, the most widespread are bar machines and single-dose pods, designed in large numbers due to their commercial popularity. This experiment tries to compare the foaming rate (Y, in percentage) effects of three different brewing techniques on espresso quality (method 1 = bar machine (BM), method 2 = hyper-espresso method (HIP), and method 3 = I-espresso system (IT)). Nine replicates per method were carried out (Table 6.53).

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for the experimental design of this study.
(b) Describe the generalized linear mixed model in (a), assuming a beta distribution.
(c) Implement the analysis of these data according to the predictor and model in (b). Summarize the relevant results.

Table 6.53 Experimental results of espresso coffee

Method	Index	Method	Index	Method	Index
1	36.64	2	70.84	3	56.19
1	39.65	2	46.68	3	36.67
1	37.74	2	73.19	3	35.35
1	35.96	2	57.78	3	40.11
1	38.52	2	48.61	3	33.52
1	21.02	2	72.77	3	37.12
1	24.81	2	65.04	3	37.33
1	34.18	2	62.53	3	32.68
1	23.08	2	54.26	3	48.33

Table 6.54 Results of wheat germination experiment in pots. Number of seeds that did not germinate out of 50

Treatments							
	1	2	3	4	5	6	7
A	10	11	8	9	7	6	9
B	8	10	3	7	9	3	11
C	5	11	2	8	10	7	11
D	1	6	4	13	7	10	10

Exercise 6.8.6 The decision to adopt a particular scale for data involving small integers is not an easy one because any analysis must be – to some extent – as adequate as possible to obtain estimates with as little uncertainty as possible. As a simple example of this type of data, consider the following results from a potted wheat germination experiment (Table 6.54).

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for this experiment.
(b) List all components of the GLMM in (a), assuming a binomial response variable.
(c) Analyze this dataset and summarize the relevant results.
(d) Is there an extra-variation in the dataset? If so, reanalyze the data with an alternative distribution. Summarize and compare your findings.

Exercise 6.8.7 A greenhouse experiment was carried out to investigate how a disease spreads in two varieties of (agurkesyge) cucumber, which is supposed to depend on the climate and the amount of fertilizers used for the two varieties. The following data come from the Department of Plant Pathology. Two climates were used: (1) change to day temperature 3 hours before sunrise and (2) normal change to day temperature. Three amounts of fertilizer were applied, normal (2.0 units), high (3.5 units), and very high (4.0 units). The two varieties were Aminex and Dalibor. To have a better controlled experiment, the plants were "standardized" to equally have as many leaves, and, then (on day 0, for example), the plants were contaminated with the disease. Subsequently, 8 days after the plants were contaminated, the amount of infection (in percentage) was recorded. From the resulting infection curve, two measures were calculated (in a manner not specified here), namely, the rate of spread of the disease (%) and the level of infection at the

end of the disease period. The experiment was implemented in three blocks, each of which consisted of two sections. Each section consisted of three plots, which were divided into two subplots, each of which had six to eight plants. Thus, there were a total of 36 subplots. The results were recorded for each subplot. The experimental factors were randomly assigned to the different units as follows: two climates to the two sections within each block, three amounts of fertilizer to the three plots within each section, and, finally, the two varieties to the two subplots within each plot. The data are shown below (Table 6.55).

(a) Write down a statistical model of this experiment.
(b) List all the components of the GLMM in (a).
(c) Write down the null and alternative hypotheses associated with this experiment.
(d) Construct an ANOVA table indicating the sources of variation and degrees of freedom.
(e) Analyze the rate of disease spread to investigate the effect of different factors.
(f) Comment on the results obtained.

Exercise 6.8.8 This example is an experiment to identify damage to the uterus in laboratory rodents after exposure to boric acid, a compound widely used in pesticides, pharmaceuticals, and other household products (Heindel et al. 1992). The study design included four doses of boric acid. The compound was administered to pregnant female mice during the first 17 days of gestation, and, then, the females were sacrificed and their litters examined. The table below presents the resulting trials for litters dying in utero (Y) of the total number of trials conducted (N) at each of the four doses tested: $d_1 = 0\{\text{control}\}$, $d_2 = 0.1$, $d_3 = 0.2$, and $d_3 = 0.4$ (as percentage of boric acid in the diet) (Table 6.56).

(a) Write down an ANOVA table (sources of variation, degrees of freedom) for this experiment.
(b) List all the components of the GLMM in (a).
(c) Analyze this dataset and summarize the relevant results.
(d) Is there an extra-variation in the dataset? If so, what alternative distribution do you propose? Reanalyze the data and compare your findings.

Table 6.55 Greenhouse experiment results of cucumber varieties

Block	Section	Plot	Weather	Fertilizer	Variety	Proportion (%)	Level
1	1	1	2	2	Aminex	48.8981	0.06915
1	1	1	2	2	Dalibor	42.2463	0.06595
1	1	2	2	3.5	Aminex	48.2108	0.04679
1	1	2	2	3.5	Dalibor	41.6767	0.04881
1	1	3	2	4	Aminex	55.4369	0.04025
1	1	3	2	4	Dalibor	40.9562	0.04859
1	2	4	1	2	Aminex	51.5573	0.09353
1	2	4	1	2	Dalibor	36.7739	0.10353
1	2	5	1	3.5	Aminex	47.9937	0.05327
1	2	5	1	3.5	Dalibor	47.8723	0.04397
1	2	6	1	4	Aminex	57.9171	0.05225
1	2	6	1	4	Dalibor	37.7185	0.09324
1	3	7	2	2	Aminex	60.1747	0.04182
2	3	7	2	2	Dalibor	45.6937	0.06983
2	3	8	2	3.5	Aminex	51.0017	0.08863
2	3	8	2	3.5	Dalibor	52.2796	0.03622
2	3	9	2	4	Aminex	51.1251	0.05875
2	3	9	2	4	Dalibor	48.7217	0.08169
2	4	10	1	2	Aminex	51.6001	0.07001
2	4	10	1	2	Dalibor	50.4463	0.09907
2	4	11	1	3.5	Aminex	48.3387	0.05788
2	4	11	1	3.5	Dalibor	38.6538	0.06834
2	4	12	1	4	Aminex	51.3147	0.05695
2	4	12	1	4	Dalibor	38.2488	0.07908
3	5	13	1	2	Aminex	49.6958	0.07218
3	5	13	1	2	Dalibor	29.6786	0.11351
3	5	14	1	3.5	Aminex	46.6692	0.08825
3	5	14	1	3.5	Dalibor	36.5892	0.09107
3	5	15	1	4	Aminex	56.032	0.04532
3	5	15	1	4	Dalibor	36.0955	0.08712
3	6	16	2	2	Aminex	45.979	0.08882
3	6	16	2	2	Dalibor	37.2489	0.12796
3	6	17	2	3.5	Aminex	40.7277	0.06418
3	6	17	2	3.5	Dalibor	38.4831	0.0854
3	6	18	2	4	Aminex	44.5242	0.06215
3	6	18	2	4	Dalibor	34.3907	0.09651

Table 6.56 Rodent experiment results

Dose	*Y*	*N*	Dose	*Y*	*N*	Dose	*Y*	*N*	Dose	*Y*	*N*
0	0	15	0.1	0	6	0.2	1	12	0.4	12	12
0	0	3	0.1	1	14	0.2	0	12	0.4	1	12
0	1	9	0.1	1	12	0.2	0	11	0.4	0	13
0	1	12	0.1	0	10	0.2	0	13	0.4	2	8
0	1	13	0.1	2	14	0.2	0	12	0.4	2	12
0	2	13	0.1	0	12	0.2	0	14	0.4	4	13
0	0	16	0.1	0	14	0.2	4	15	0.4	0	13
0	0	11	0.1	3	14	0.2	0	14	0.4	1	13
0	1	11	0.1	0	10	0.2	0	12	0.4	0	12
0	2	8	0.1	2	12	0.2	1	6	0.4	1	9
0	0	14	0.1	3	13	0.2	2	13	0.4	3	9
0	0	13	0.1	1	11	0.2	0	10	0.4	0	11
0	3	14	0.1	1	11	0.2	1	14	0.4	1	14
0	1	13	0.1	0	11	0.2	1	12	0.4	0	10
0	0	8	0.1	0	13	0.2	0	10	0.4	3	12
0	0	13	0.1	0	10	0.2	0	9	0.4	2	21
0	2	14	0.1	1	12	0.2	1	12	0.4	3	10
0	3	14	0.1	0	11	0.2	0	13	0.4	3	11
0	0	11	0.1	2	10	0.2	1	14	0.4	1	11
0	2	12	0.1	2	12	0.2	0	13	0.4	1	11
0	0	15	0.1	2	15	0.2	0	14	0.4	8	14
0	0	15	0.1	3	12	0.2	1	13	0.4	0	15
0	2	14	0.1	1	12	0.2	2	12	0.4	2	13
0	1	11	0.1	0	12	0.2	1	14	0.4	8	11
0	1	16	0.1	1	12	0.2	0	13	0.4	4	12
0	0	12	0.1	1	13	0.2	0	12	0.4	2	12
0	0	14	0.1	1	15	0.2	1	7			

Appendix

Data: Fleas					
Bioen	SP	Treat	Rep	Overvi	Dead
B1	Daphnia	T1	1	10	0
B1	Daphnia	T1	2	10	0
B1	Daphnia	T1	3	10	0
B1	Daphnia	T2	1	10	0
B1	Daphnia	T2	2	10	0
B1	Daphnia	T2	3	10	0
B1	Daphnia	T3	1	9	1
B1	Daphnia	T3	2	9	1
B1	Daphnia	T3	3	8	2

(continued)

Data: Fleas

Bioen	SP	Treat	Rep	Overvi	Dead
B1	Daphnia	T4	1	2	8
B1	Daphnia	T4	2	2	8
B1	Daphnia	T4	3	3	7
B1	Daphnia	T5	1	0	10
B1	Daphnia	T5	2	0	10
B1	Daphnia	T5	3	0	10
B1	Daphnia	T6	1	0	10
B1	Daphnia	T6	2	0	10
B1	Daphnia	T6	3	0	10
B2	Daphnia	T1	1	10	0
B2	Daphnia	T1	2	10	0
B2	Daphnia	T1	3	10	0
B2	Daphnia	T2	1	10	0
B2	Daphnia	T2	2	10	0
B2	Daphnia	T2	3	10	0
B2	Daphnia	T3	1	9	1
B2	Daphnia	T3	2	9	1
B2	Daphnia	T3	3	9	1
B2	Daphnia	T4	1	2	8
B2	Daphnia	T4	2	2	8
B2	Daphnia	T4	3	2	8
B2	Daphnia	T5	1	0	10
B2	Daphnia	T5	2	0	10
B2	Daphnia	T5	3	0	10
B2	Daphnia	T6	1	0	10
B2	Daphnia	T6	2	0	10
B2	Daphnia	T6	3	0	10
B3	Daphnia	T1	1	10	0
B3	Daphnia	T1	2	10	0
B3	Daphnia	T1	3	10	0
B3	Daphnia	T2	1	10	0
B3	Daphnia	T2	2	10	0
B3	Daphnia	T2	3	10	0
B3	Daphnia	T3	1	8	2
B3	Daphnia	T3	2	9	1
B3	Daphnia	T3	3	9	1
B3	Daphnia	T4	1	3	7
B3	Daphnia	T4	2	2	8
B3	Daphnia	T4	3	2	8
B3	Daphnia	T5	1	0	10
B3	Daphnia	T5	2	0	10
B3	Daphnia	T5	3	0	10

(continued)

Data: Fleas					
Bioen	SP	Treat	Rep	Overvi	Dead
B3	Daphnia	T6	1	0	10
B3	Daphnia	T6	2	0	10
B3	Daphnia	T6	3	0	10
B1	Dubia	T1	1	10	0
B1	Dubia	T1	2	10	0
B1	Dubia	T1	3	10	0
B1	Dubia	T2	1	5	5
B1	Dubia	T2	2	6	4
B1	Dubia	T2	3	6	4
B1	Dubia	T3	1	5	5
B1	Dubia	T3	2	5	5
B1	Dubia	T3	3	5	5
B1	Dubia	T4	1	2	8
B1	Dubia	T4	2	3	7
B1	Dubia	T4	3	3	7
B1	Dubia	T5	1	2	8
B1	Dubia	T5	2	2	8
B1	Dubia	T5	3	2	8
B1	Dubia	T6	1	0	10
B1	Dubia	T6	2	0	10
B1	Dubia	T6	3	0	10
B2	Dubia	T1	1	10	0
B2	Dubia	T1	2	10	0
B2	Dubia	T1	3	10	0
B2	Dubia	T2	1	7	3
B2	Dubia	T2	2	5	5
B2	Dubia	T2	3	6	4
B2	Dubia	T3	1	5	5
B2	Dubia	T3	2	5	5
B2	Dubia	T3	3	5	5
B2	Dubia	T4	1	4	6
B2	Dubia	T4	2	4	6
B2	Dubia	T4	3	4	6
B2	Dubia	T5	1	2	8
B2	Dubia	T5	2	2	8
B2	Dubia	T5	3	2	8
B2	Dubia	T6	1	0	10
B2	Dubia	T6	2	0	10
B2	Dubia	T6	3	0	10
B3	Dubia	T1	1	10	0
B3	Dubia	T1	2	10	0
B3	Dubia	T1	3	10	0

(continued)

Data: Fleas

Bioen	SP	Treat	Rep	Overvi	Dead
B3	Dubia	T2	1	8	2
B3	Dubia	T2	2	8	2
B3	Dubia	T2	3	7	3
B3	Dubia	T3	1	5	5
B3	Dubia	T3	2	5	5
B3	Dubia	T3	3	6	4
B3	Dubia	T4	1	2	8
B3	Dubia	T4	2	3	7
B3	Dubia	T4	3	2	8
B3	Dubia	T5	1	3	7
B3	Dubia	T5	2	2	8
B3	Dubia	T5	3	2	8
B3	Dubia	T6	1	0	10
B3	Dubia	T6	2	0	10
B3	Dubia	T6	3	0	10

Data: Commercial crop explant detachment

Block	A	B	C	y	N
1	1	1	1	15	73
2	1	1	1	10	86
1	1	1	2	17	69
2	1	1	2	19	32
1	1	2	1	26	125
2	1	2	1	21	62
1	1	2	2	14	81
2	1	2	2	12	21
1	2	1	1	10	92
2	2	1	1	12	108
1	2	1	2	30	44
2	2	1	2	32	33
1	2	2	1	37	91
2	2	2	1	30	42
1	2	2	2	32	98
2	2	2	2	37	44
1	3	1	1	18	52
2	3	1	1	18	73
1	3	1	2	23	108
2	3	1	2	21	55
1	3	2	1	24	106
2	3	2	1	27	92
1	3	2	2	37	64
2	3	2	2	37	97

Data: Cockroaches (E1 = np, E2 = ng, E3 = adult)			
Bioassay	Isolation	Age	Dead
1	Bb1	np	7
2	Bb1	np	6
1	Bb1	ng	2
2	Bb1	ng	2
1	Bb1	a	9
2	Bb1	a	6
1	Bb2	np	6
2	Bb2	np	7
1	Bb2	ng	7
2	Bb2	ng	3
1	Bb2	a	10
2	Bb2	a	8
1	Bb3	np	3
2	Bb3	np	2
1	Bb3	ng	2
2	Bb3	ng	3
1	Bb3	a	8
2	Bb3	a	9
1	Bb4	np	6
2	Bb4	np	4
1	Bb4	ng	5
2	Bb4	ng	3
1	Bb4	a	10
2	Bb4	a	8
1	Bb5	np	7
2	Bb5	np	7
1	Bb5	ng	3
2	Bb5	ng	1
1	Bb5	a	7
2	Bb5	a	3
1	Bb6	np	7
2	Bb6	np	9
1	Bb6	ng	6
2	Bb6	ng	2
1	Bb6	a	10
2	Bb6	a	7
1	Bb8	np	9
2	Bb8	np	9
1	Bb8	ng	4
2	Bb8	ng	2
1	Bb8	a	9
2	Bb8	a	10

(continued)

Data: Cockroaches (E1 = np, E2 = ng, E3 = adult)			
Bioassay	Isolation	Age	Dead
1	Bb9	np	5
2	Bb9	np	8
1	Bb9	ng	6
2	Bb9	ng	2
1	Bb9	a	7
2	Bb9	a	5
1	Bb10	np	8
2	Bb10	np	6
1	Bb10	ng	1
2	Bb10	ng	4
1	Bb10	a	3
2	Bb10	a	4
1	Bb11	np	8
2	Bb11	np	7
1	Bb11	ng	1
2	Bb11	ng	3
1	Bb11	a	6
2	Bb11	a	8
1	Bb12	np	8
2	Bb12	np	9
1	Bb12	ng	8
2	Bb12	ng	9
1	Bb12	a	7
2	Bb12	a	6
1	Bb13	np	6
2	Bb13	np	3
1	Bb13	ng	0
2	Bb13	ng	1
1	Bb13	a	5
2	Bb13	a	6
1	Bb14	np	10
2	Bb14	np	5
1	Bb14	ng	4
2	Bb14	ng	2
1	Bb14	a	6
2	Bb14	a	6
1	Bb15	np	5
2	Bb15	np	10
1	Bb15	ng	6
2	Bb15	ng	1
1	Bb15	a	4
2	Bb15	a	5

(continued)

Data: Cockroaches (E1 = np, E2 = ng, E3 = adult)			
Bioassay	Isolation	Age	Dead
1	Bb16	np	5
2	Bb16	np	7
1	Bb16	ng	3
2	Bb16	ng	4
1	Bb16	a	8
2	Bb16	a	6
1	Control	np	1
2	Control	np	0
1	Control	ng	0
2	Control	ng	0
1	Control	a	0
2	Control	a	1

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
1	1	1	1	1	14
1	1	1	2	2	12
1	1	1	3	0	12
1	1	1	4	0	13
1	1	1	5	3	8
1	1	1	6	0	9
1	1	2	1	7	8
1	1	2	2	0	10
1	1	2	3	1	14
1	1	2	4	0	10
1	1	2	5	0	17
1	1	2	6	0	10
1	1	3	1	9	14
1	1	3	2	1	11
1	1	3	3	0	10
1	1	3	4	1	14
1	1	3	5	0	10
1	1	3	6	0	21
1	1	4	1	10	17
1	1	4	2	0	9
1	1	4	3	1	12
1	1	4	4	0	11
1	1	4	5	0	12
1	1	4	6	0	10
1	1	5	1	8	11
1	1	5	2	1	10

(continued)

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
1	1	5	3	0	9
1	1	5	4	2	12
1	1	5	5	0	10
1	1	5	6	1	11
1	2	1	1	7	9
1	2	1	2	2	10
1	2	1	3	0	10
1	2	1	4	0	14
1	2	1	5	1	12
1	2	1	6	0	13
1	2	2	1	6	12
1	2	2	2	0	11
1	2	2	3	1	13
1	2	2	4	0	9
1	2	2	5	2	11
1	2	2	6	0	10
1	2	3	1	6	7
1	2	3	2	1	12
1	2	3	3	0	9
1	2	3	4	1	10
1	2	3	5	0	14
1	2	3	6	2	12
1	2	4	1	7	13
1	2	4	2	0	10
1	2	4	3	0	10
1	2	4	4	1	12
1	2	4	5	0	9
1	2	4	6	1	8
1	2	5	1	11	15
1	2	5	2	1	13
1	2	5	3	0	14
1	2	5	4	1	14
1	2	5	5	0	11
1	2	5	6	0	11
1	3	1	1	5	11
1	3	1	2	5	11
1	3	1	3	0	15
1	3	1	4	1	15
1	3	1	5	0	8
1	3	1	6	1	10
1	3	2	1	4	9
1	3	2	2	1	15

(continued)

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
1	3	2	3	0	11
1	3	2	4	0	13
1	3	2	5	1	12
1	3	2	6	0	12
1	3	3	1	9	12
1	3	3	2	2	14
1	3	3	3	0	12
1	3	3	4	0	12
1	3	3	5	0	10
1	3	3	6	0	13
1	3	4	1	10	10
1	3	4	2	0	10
1	3	4	3	0	8
1	3	4	4	0	10
1	3	4	5	1	14
1	3	4	6	3	11
1	3	5	1	9	11
1	3	5	2	0	11
1	3	5	3	1	11
1	3	5	4	1	14
1	3	5	5	0	9
1	3	5	6	0	9
2	1	1	1	0	12
2	1	1	2	0	12
2	1	1	3	0	14
2	1	1	4	0	12
2	1	1	5	0	10
2	1	1	6	1	13
2	1	2	1	8	9
2	1	2	2	1	9
2	1	2	3	0	12
2	1	2	4	0	10
2	1	2	5	0	12
2	1	2	6	1	10
2	1	3	1	11	14
2	1	3	2	1	12
2	1	3	3	1	11
2	1	3	4	0	10
2	1	3	5	0	9
2	1	3	6	3	11
2	1	4	1	12	15
2	1	4	2	0	15

(continued)

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
2	1	4	3	0	10
2	1	4	4	1	9
2	1	4	5	1	10
2	1	4	6	0	16
2	1	5	1	10	14
2	1	5	2	1	9
2	1	5	3	0	11
2	1	5	4	0	11
2	1	5	5	0	11
2	1	5	6	0	11
2	2	1	1	1	9
2	2	1	2	0	9
2	2	1	3	0	12
2	2	1	4	1	10
2	2	1	5	1	12
2	2	1	6	0	17
2	2	2	1	9	12
2	2	2	2	0	12
2	2	2	3	0	11
2	2	2	4	2	14
2	2	2	5	0	11
2	2	2	6	0	10
2	2	3	1	7	13
2	2	3	2	0	16
2	2	3	3	1	12
2	2	3	4	0	10
2	2	3	5	0	10
2	2	3	6	0	11
2	2	4	1	7	13
2	2	4	2	1	18
2	2	4	3	0	10
2	2	4	4	0	11
2	2	4	5	0	11
2	2	4	6	3	13
2	2	5	1	5	10
2	2	5	2	0	10
2	2	5	3	0	10
2	2	5	4	0	10
2	2	5	5	0	9
2	2	5	6	1	12
2	3	1	1	6	13
2	3	1	2	0	10

(continued)

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
2	3	1	3	1	11
2	3	1	4	3	11
2	3	1	5	0	12
2	3	1	6	1	19
2	3	2	1	12	13
2	3	2	2	0	11
2	3	2	3	0	8
2	3	2	4	0	9
2	3	2	5	0	17
2	3	2	6	0	12
2	3	3	1	8	11
2	3	3	2	4	12
2	3	3	3	0	11
2	3	3	4	0	10
2	3	3	5	0	15
2	3	3	6	3	13
2	3	4	1	5	14
2	3	4	2	1	9
2	3	4	3	0	12
2	3	4	4	1	12
2	3	4	5	0	10
2	3	4	6	2	14
2	3	5	1	10	14
2	3	5	2	0	14
2	3	5	3	1	10
2	3	5	4	1	13
2	3	5	5	1	15
2	3	5	6	4	10
3	1	1	1	8	12
3	1	1	2	1	14
3	1	1	3	0	12
3	1	1	4	0	20
3	1	1	5	1	18
3	1	1	6	7	15
3	1	2	1	9	16
3	1	2	2	1	12
3	1	2	3	0	13
3	1	2	4	0	15
3	1	2	5	0	17
3	1	2	6	1	18
3	1	3	1	7	12
3	1	3	2	0	14

(continued)

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
3	1	3	3	1	13
3	1	3	4	0	18
3	1	3	5	0	14
3	1	3	6	0	14
3	1	4	1	10	14
3	1	4	2	2	17
3	1	4	3	0	10
3	1	4	4	1	19
3	1	4	5	0	17
3	1	4	6	0	16
3	1	5	1	9	10
3	1	5	2	1	14
3	1	5	3	1	11
3	1	5	4	0	18
3	1	5	5	0	15
3	1	5	6	1	11
3	2	1	1	10	10
3	2	1	2	1	11
3	2	1	3	0	12
3	2	1	4	1	15
3	2	1	5	4	20
3	2	1	6	0	14
3	2	2	1	9	12
3	2	2	2	1	10
3	2	2	3	1	12
3	2	2	4	3	18
3	2	2	5	0	16
3	2	2	6	0	12
3	2	3	1	10	11
3	2	3	2	1	16
3	2	3	3	1	14
3	2	3	4	1	17
3	2	3	5	2	15
3	2	3	6	1	16
3	2	4	1	9	11
3	2	4	2	2	14
3	2	4	3	0	10
3	2	4	4	0	18
3	2	4	5	0	17
3	2	4	6	0	12
3	2	5	1	11	12
3	2	5	2	2	12

(continued)

Data: Disease incidence in grapevine plants (b = block, v = plant, r = shoot, t = treatment, m = number of diseased leaves per shoot, and n = total number of leaves per shoot).

b	v	r	t	M	n
3	2	5	3	0	11
3	2	5	4	0	13
3	2	5	5	0	18
3	2	5	6	0	12
3	3	1	1	7	9
3	3	1	2	0	13
3	3	1	3	0	9
3	3	1	4	0	18
3	3	1	5	0	18
3	3	1	6	0	13
3	3	2	1	6	14
3	3	2	2	3	16
3	3	2	3	1	15
3	3	2	4	0	17
3	3	2	5	1	17
3	3	2	6	3	14
3	3	3	1	10	11
3	3	3	2	0	10
3	3	3	3	1	16
3	3	3	4	1	18
3	3	3	5	0	16
3	3	3	6	0	11
3	3	4	1	10	10
3	3	4	2	1	14
3	3	4	3	0	10
3	3	4	4	1	19
3	3	4	5	2	19
3	3	4	6	2	14
3	3	5	1	8	10
3	3	5	2	0	12
3	3	5	3	0	12
3	3	5	4	0	18
3	3	5	5	0	14
3	3	5	6	0	12