

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326807815>

Overdispersed fungus germination data: statistical analysis using R

Article in *Biocontrol Science and Technology* · August 2018

DOI: 10.1080/09583157.2018.1504888

CITATIONS

5

READS

225

8 authors, including:



Maira Fatoretto

University of São Paulo

3 PUBLICATIONS 5 CITATIONS

[SEE PROFILE](#)



Rafael De Andrade Moral

National University of Ireland, Maynooth

56 PUBLICATIONS 294 CITATIONS

[SEE PROFILE](#)



Clarice G. B. Demétrio

University of São Paulo

179 PUBLICATIONS 3,056 CITATIONS

[SEE PROFILE](#)



Víctor Manuel Arevalo Rojas

Universidad Nacional de Barranca, Lima, Perú

1 PUBLICATION 5 CITATIONS

[SEE PROFILE](#)

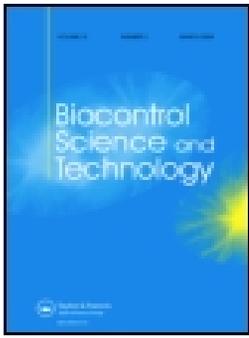
Some of the authors of this publication are also working on these related projects:



Tolerance of Eucalyptus Clone to Hydric and Thermal Stresses (TECHS) [View project](#)



Flexible regression models for count data [View project](#)



Overdispersed fungus germination data: statistical analysis using R

Maíra Blumer Faretto, Rafael de Andrade Moral, Clarice Garcia Borges Demétrio, Christopher Silva de Pádua, Vinicius Menarin, Víctor Manuel Arévalo Rojas, Celeste Paola D'Alessandro & Italo Delalibera Jr.

To cite this article: Maíra Blumer Faretto, Rafael de Andrade Moral, Clarice Garcia Borges Demétrio, Christopher Silva de Pádua, Vinicius Menarin, Víctor Manuel Arévalo Rojas, Celeste Paola D'Alessandro & Italo Delalibera Jr. (2018): Overdispersed fungus germination data: statistical analysis using R, *Biocontrol Science and Technology*, DOI: [10.1080/09583157.2018.1504888](https://doi.org/10.1080/09583157.2018.1504888)

To link to this article: <https://doi.org/10.1080/09583157.2018.1504888>



View supplementary material [↗](#)



Published online: 02 Aug 2018.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)



Overdispersed fungus germination data: statistical analysis using R

Maíra Blumer Fatoreto ^a, Rafael de Andrade Moral^b, Clarice Garcia Borges Demétrio^a, Christopher Silva de Pádua^a, Vinicius Menarin^a, Víctor Manuel Arévalo Rojas^a, Celeste Paola D'Alessandro^a and Italo Delalibera Jr.^a

^aESALQ, Universidade de São Paulo, Piracicaba, Brazil; ^bDepartment of Mathematics and Statistics, Maynooth University, Maynooth, Ireland

ABSTRACT

Proportion data from dose-response experiments are often overdispersed, characterised by a larger variance than assumed by the standard binomial model. Here, we present different models proposed in the literature that incorporate overdispersion. We also discuss how to select the best model to describe the data and present, using R software, specific code used to fit and interpret binomial, quasi-binomial, beta-binomial, and binomial-normal models, as well as to assess goodness-of-fit. We illustrate applications of these generalised linear models and generalised linear mixed models with a case study from a biological control experiment, where different isolates of *Isaria fumosorosea* (Hypocreales: Cordycipitaceae) were used to assess which ones presented higher resistance to UV-B radiation. We show how to test for differences between isolates and also how to statistically group isolates presenting a similar behaviour.

ARTICLE HISTORY

Received 8 April 2018
Accepted 20 July 2018

KEYWORDS

Entomopathogenic fungi;
generalised linear models;
mixed models; proportion
data; random effects

Introduction

The class of generalised linear models (GLM) was introduced by Nelder and Wedderburn (1972) as a general framework for handling a range of statistical models for normal and non-normal data. The GLM framework allows us to analyze proportion data.

An extension of these models are the generalised linear mixed models (GLMM), which include random effects in the linear predictor, and can be used to adequately accommodate extra variability. When the variation is greater than expected by the standard Poisson and binomial models, natural starting points for the analysis of count and proportion data, respectively, this phenomenon is referred to as overdispersion (Hinde & Demétrio, 1998). In the presence of overdispersion, an alternative is to include random effects in the linear predictor of the model to explain the extra-variability.

Germination data are usually overdispersed (Hinde & Demétrio, 1998). This may be due to individual variability of the experimental units, or due to some correlation between these units. For example, experimental units consisting of a Petri dish or a test

CONTACT Maíra Blumer Fatoreto  mairafatoreto@gmail.com  ESALQ, Universidade de São Paulo, Avenida Pádua Dias, 11, Piracicaba, SP 13418-900 Brazil

 Supplemental data for this article can be accessed <https://doi.org/10.1080/09583157.2018.1504888>.

© 2018 Informa UK Limited, trading as Taylor & Francis Group

tube that are set up using the same solutions have a higher correlation than units with different solutions. The main problem of not considering overdispersion in the model when it is present is the under estimation of standard errors and, consequently, erroneous interpretation and conclusions drawn.

One of the interests of fungus germination data is to assess the Ultraviolet radiation (UV-B) tolerance of reproductive structures (conidia) of entomopathogenic fungi. The conidia germination is valued to select the most tolerant fungal isolate that could be used for the development of biopesticides. However, some isolates may present similar resistance and, in this case, it is necessary to cluster them according the statistical results.

In this paper, we discuss how to select models to properly analyze overdispersed proportion data, using R (R Core Team, 2017). We then illustrate applications of generalised linear models and generalised linear mixed models, using fungal germination data. Finally, once the final model is selected, we present methods for grouping isolates and testing for differences between them.

Case study

An experiment in a randomised block complete design was conducted in a laboratory to study the UV-B radiation tolerance on fourteen isolates of *Isaria fumosorosea* (Wize). The entomopathogenic fungus *Isaria fumosorosea* (Wize) Brown & Smith (Ascomycota: Hypocreales: Cordycipitaceae) is commonly found in soil and infecting several species of arthropods (Zimmermann, 2008). Mycoinsecticides based on *I. fumosorosea* were developed and commercialised in the world being used to control insects of the order Hemiptera (whitefly, aphids and scales), Thysanoptera (thrips), Acari (spider mites) and Coleoptera (Faria & Wraight, 2007; Zimmermann, 2008). However, in Brazil there is only one bioproduct based on this pathogen to *Diaphorina citri* control. The objective of this experiment was to select one strain of *I. fumosorosea* with high UV-B radiation tolerance for development of a new mycopesticide based on *I. fumosorosea*. The *I. fumosorosea* strains were obtained from soil and insects of different Brazilian biomes and preserved in the entomopathogen collection of the Laboratory of Pathology and Microbial Control of Insects at the 'Luiz de Queiroz' College of Agriculture (ESALQ), University of Sao Paulo (USP), located in Piracicaba, SP, Brazil. The conidia suspension of each isolate was inoculated in a one Petri dish containing culture medium and all Petri dishes were placed to a wooden chamber with four fluorescent lamps emitting UV-B light. Irradiation experiments were conducted in a chamber with four UV-B 313EL lamps (Q-lab Cleveland, OH, USA). The lamps were aged prior to the start the experiments, resulting in a stable level of irradiation of 1312.4 mW/m² and 5.20 KJ/h. Therefore, the exposure times were 0, 2, 4, 6 and 8 h, which correspond to total UV-B irradiation of 0, 4.76, 9.52, 14.28 and 19.04 KJ/m², respectively. The Petri dishes were covered with a 0.13 mm-thick cellulose diacetate film (Málaga Ltda), which had a cutoff point at 290 nm. This permitted the passage of most UV-B and UV-A (290–360 nm), with was selected based on the spectral characteristics of fungal responses reviewed by Paul, Rasanayagam, Moody, Hatcher, and Ayres (1997). After exposure to UV-B light, the Petri dishes were maintained in a growth chamber set to 26 ± 2°C under a 12 h photophase for 48 h to allow the recovery and germination of conidia. The control treatment consisted of Petri dishes containing the conidia suspension of each isolate without exposure to UV-B radiation. Germination of

this treatment was quantified after 24 h of incubation in a growth chamber at 26°C and 12 h of photophase. The direct count of germinated and non-germinated conidia was performed under a light microscope following the methodology of Oliveira, Pauli, Mascarin, and Delalibera (2015). After that, the dishes were incubated for 48 h to allow for fungi recovery and conidia germination. The direct counting of the number of germinated and non-germinated conidia was carried out using an optical microscope (400x), and conidia were considered as germinated if their germ tube was twice as long as the spore. All treatments were repeated four times in the same conditions. In each of the replicates (blocks), was prepared one conidia suspension of each isolate and used for inoculation in five Petri dishes corresponding to each exposure time (0, 2, 4, 6 and 8 h), adding up to 14 isolates \times 5 exposure time (hours) \times 4 blocks (replicates) = 280 observations. In this case, we may expect correlation between the isolates per block. There is also an inherent variability of the application of fungal suspensions in the centre of the Petri dishes due to the use of hand-held instruments (micropipette), or conduction by different people. We present the observed means and standard errors for each combination between isolate and exposure time as online Supplementary Material.

The exploratory plots of the data (Figure 1) show the proportions of germinated conidia for each fungus, over time of exposure. At the beginning of the experiment, the isolate is less exposed to UV-B radiation, thus, has a higher germination rate. However, the longer the exposure time, greater is the damage to the isolate, which causes the germination reduction and, in some cases, the death of the conidia. There is large variability between blocks within isolates (Figure 1(a)) and also between isolates (Figure 1(b)). Also, there are different starting points for each curve at time 0 (Figure 1(b)), i.e. the isolates have different initial proportions of conidia germination.

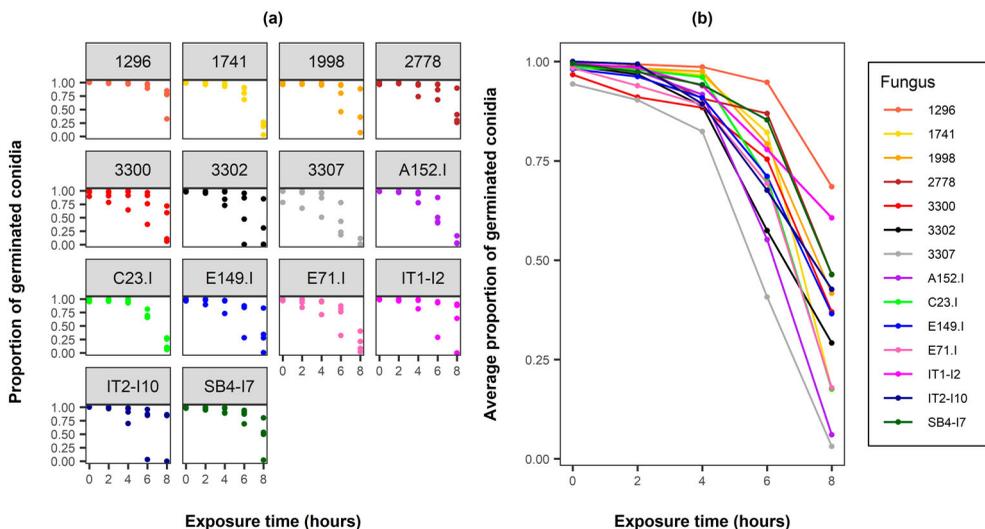


Figure 1. Plots of (a) observed proportion of germinated conidia vs. exposure time per isolate and (b) averages of observed proportion of germinated conidia vs. exposure time for each isolate.

Statistical methods

Here, the outcome of interest is the proportion of germinated conidia, hence the response variable is the ratio between the number of germinated conidia, represented by the random variables Y_i , $i = 1, \dots, n$, and the total number of conidia in the i -th Petri dish, m_i , $i = 1, \dots, n$.

If we assume each conidia has the same germination probability, then we can assume the binomial distribution for our variables of interest Y_i . The binomial model is a reasonable starting point to analyze these data, and extensions are available, such as the ones discussed in latter sections. This is a specific example of a generalised linear model (Nelder & Wedderburn, 1972).

Introduction to generalized linear models (GLMs)

The class of GLMs consists of three components: (i) the random component of the model, i.e. a univariate distribution belonging to the exponential family, with probability density function (pdf)

$$f(y_i; \theta_i; \phi) = \exp\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\} \quad i = 1, \dots, n; \quad (1)$$

where $b(\cdot)$ and $c(\cdot)$ are known functions, θ_i is called the natural parameter and ϕ the dispersion parameter (for the normal distribution, for example, we have $\theta = \mu$ and $\phi = \sigma^2$, the mean and variance, respectively); (ii) a linear predictor related to the explanatory variables

$$\eta_i = \boldsymbol{\beta}^T \mathbf{x}_i$$

where $\boldsymbol{\beta}$ is a vector of p unknown parameters and $\mathbf{x}_i = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ is the i -th column of the $n \times p$ design matrix; and (iii) a link function $g(\mu_i) = \eta_i$, relating the systematic to the random component (Demétrio, Hinde, & Moral, 2014). We have that the mean is $E[Y_i] = \mu_i$ and the variance is $\text{Var}[Y_i] = \phi b''(\theta_i) = \phi V(\mu_i)$, where $V(\mu_i)$ is called variance function.

Nelder and Wedderburn (1972) proposed the analysis of deviance, generalising the ideas in standard ANOVA, to assess the significance of effects in the linear predictor. In this framework, a measure that compares a fitted model to the saturated model (i.e. a model that has one parameter per observation), and, for known ϕ , can be used as a measurement of goodness-of-fit for the fitted model, is the residual deviance. For the binomial model, the residual deviance can be written as

$$D_B = 2 \sum_{i=1}^n \left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) + (m_i - y_i) \log\left(\frac{m_i - y_i}{m_i - \widehat{\mu}_i}\right) \right]$$

where $\widehat{\mu}_i$, $i = 1, 2, \dots, n$, are the fitted values for the current model. Asymptotically (i.e. for a large sample size), D_B has an approximate χ^2 distribution with $n - p$ degrees of freedom (df).

To compare nested models, which are models that contain the same terms and one has at least one additional term, by writing D_p for the residual deviance of the full model and D_q for the reduced model, the statistic $D_q - D_p \sim \chi_{q-p}^2$ can be used to test the hypothesis

that true coefficient values of omitted terms are zero. This test corresponds to a likelihood-ratio test, and if $D_q - D_p > \chi_{q-p; (1-\alpha)}^2$, the upper $100 \times \alpha$ percentile of the χ_{q-p}^2 distribution, we reject the null hypothesis that the additional parameter is zero at a significance level of α , which means that the parameters tested are important to describe the data and should remain in the model (Demétrio et al., 2014).

To detect possible model failure we can use diagnostic plots. In general, we compare observed and fitted values using some type of residuals, usually the deviance residuals. Atkinson (1985) proposed the addition of a simulated envelope to half-normal plots to check the goodness-of-fit of a given model to a set of observations. The half-normal plots show the residuals (or any other model diagnostic) should behave if the observed data were a plausible realisation of the fitted model, making it possible to detect overdispersion in the data. If many points are outside the envelope, then the fitted model is not suitable for analyzing the data. A fit is usually considered satisfactory when the number of points outside of the simulation envelope falls below 5%. In R, these plots are implemented as the `hnp` package (Moral, Hinde, & Demétrio, 2017).

Binomial model

Let Y_i be a random variable representing counts of successes out of a sample of size m_i , $i = 1, \dots, n$, then a reasonable assumption would be that $Y_i \sim \text{Bin}(m_i, \pi_i)$, with probability function (pf)

$$P[Y_i = y_i] = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i}, y_i = 0, 1, \dots, m_i.$$

We have that $E[Y_i] = \mu_i = m_i \pi_i$ and

$$\text{Var}[Y_i] = m_i \pi_i (1 - \pi_i) \quad (2)$$

The GLM framework allows us to model the expected proportions π_i in terms of explanatory variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ (Hinde & Demétrio, 1998). The usual (canonical) link function for the binomial GLM is the logit

$$g(\mu_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (3)$$

This link function maps the probabilities π_i to the real line, and yields the characteristic sigmoid curve used to describe the behaviour of the response variable in dose-response experiments.

Because π_i is bounded between 0 and 1, $\text{Var}[Y_i] = (1 - \pi_i)E[Y_i] < E[Y_i]$ and hence, when the variability of the data is larger than the mean, this model is not able to accommodate overdispersion. Therefore, extensions of the binomial model can be used to analyze overdispersed data, such as the ones described in the following sections.

Quasi-binomial model

The simplest way to deal with overdispersion is based on a quasi-likelihood approach, which requires the specification of the first two moments of the distribution, i.e. the

mean and the variance. For the binomial GLM, Equation (2) is replaced by

$$\text{Var}[Y_i] = \phi m_i \pi_i (1 - \pi_i), \quad (4)$$

where ϕ is called the dispersion parameter. Differently from the binomial model, in which ϕ is known ($\phi = 1$), here it is considered as an unknown parameter and a way of estimating ϕ is through

$$\tilde{\phi} = \frac{X_B^2}{n - p}$$

where $X_B^2 = \sum_{i=1}^n (y_i - m_i \hat{\pi}_i)^2 / (m_i \hat{\pi}_i (1 - \hat{\pi}_i))$ is the generalised Pearson statistic for the binomial model, a measure of goodness-of-fit, such as the residual deviance.

Beta-binomial model

One of the assumptions of the binomial model is that the probability of success is the same for every Bernoulli trial.¹ An alternative is to allow the probability of success to vary according to some distribution. Because this probability is continuous and can assume values in the interval (0,1), the beta distribution is a suitable option. Let $Y_i | P_i \sim \text{Binomial}(m_i, P_i)$, with $P_i \sim \text{Beta}(a_i, b_i)$. Then marginally, Y_i has a beta-binomial distribution with mean and variance given by

$$E[Y_i] = m_i \pi_i \quad (5)$$

and

$$\text{Var}[Y_i] = m_i \pi_i (1 - \pi_i) [1 + \phi(m_i - 1)], \quad (6)$$

where $\pi_i = a_i / (a_i + b_i)$ and $\phi = 1 / (a_i + b_i + 1)$ with $(a_i + b_i)$ considered to be constant. This model is often used when the binomial denominator m_i is variable (Demétrio et al., 2014).

Hence, if $\phi = 0$ the variance function in Equation (6) will be of the same form as in the binomial model (Equation (2)). If the m_i are equal, the variance function in Equation (6) will be of the same form as in the quasi-binomial model (Equation (4)).

Logistic-normal model

Another way to incorporate overdispersion is to assume that there is random variation in the linear predictor η_i . This can be done by including a random effect in the linear predictor, i.e. $\eta_i = x_i' \beta + Z_i$, with $Z_i \sim N(0, \sigma^2)$. Using a logit link (Equation (3)), we obtain the logistic-normal model. The expectation of Y_i is the same as in Equation (5), and the variance may be approximated by

$$\text{Var}[Y_i] \approx m_i \pi_i (1 - \pi_i) [1 + \sigma^2 (m_i - 1) \pi_i (1 - \pi_i)]$$

This is the simplest case of a GLMM, because it includes fixed and random effects in the linear predictor.

In some applications these random effects can be interpreted, e.g. when there is random variation in the conidia concentration of the solutions, this may increase data variability,

and observational-level random effects can be included in the linear predictor to explain this extra-variability.

Here, we use likelihood-ratio tests to assess the significance of the variance components. The null hypothesis for a variance component σ^2 is $H_0: \sigma^2 = 0$ versus the alternative $H_1: \sigma^2 > 0$. Because this test is on the boundary of the parameter space (i.e. the

variance cannot be negative), the resulting asymptotic distribution of the test statistic is a mixture of chi-square distributions. For more details, see Verbeke and Molenberghs (2003).

Analysis the case-study

Fitting the binomial model

We begin by fitting a standard binomial logit model (M1), using the maximal linear predictor for this case, i.e. with different logistic regression curves (one for each of the different isolates, with different intercept by block):

$$\eta_{jik} = \gamma_j + \beta_{0i} + \beta_{1i}t_k, \quad j = 1, \dots, 4, \quad i = 1, \dots, 14 \quad \text{and} \quad k = 1, \dots, 5, \quad (7)$$

where γ_j is the effect of j -th block, β_{0i} and β_{1i} are the intercept and slope for the i -th isolate, respectively and t_k is the k -th exposure time. Here, the outcome variable is the proportion of germinated conidia and we have two covariates: the identification of the fungus isolate, represented as a categorical variable in the model, and the dose of exposure time represented in the model as a continuous variable. The R code to carry out the analysis is given in [Box 1](#).

Box 1. R code to fit the standard binomial model to the germination data.

```
# read data
germination <- read.table(' data.txt', header=TRUE, dec='.', sep=' \t' )
# examining first 6 rows of the dataset
head(germination)
FUNG    DOSE    BLOCK    TOTAL    GERM
1 1296     0     1      126     126
2 1296     2     1      112     111
3 1296     4     1      110     108
4 1296     6     1       68     66
5 1296     8     1       74     24
6 1296     0     2       88     88
resp <- with(dados, cbind(GERM, TOTAL-GERM))
# fitting binomial model (M1) using block and fung as factor,
# dose as numeric and the linear predictor (7)
M1 <- glm(resp ~ BLOCK+FUNG*DOSE, family=binomial,
data=germination)
# check adequacy of model using hnp
require(hnp)
```

Looking at the half-normal plot with simulated envelope for the deviance residuals ([Figure 2\(a\)](#)), there is evidence that the model does not fit the data satisfactorily. We are looking here for clear patterns of systematic deviations from the expected shape of the envelope, which will mean that the observed values are not a plausible realisation of the fitted model and thus it should not be used to analyze the data. Here, the reason for

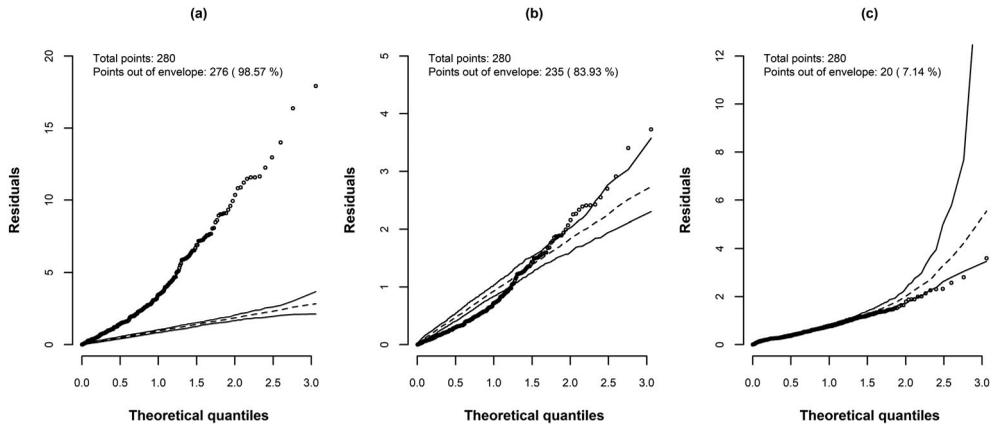


Figure 2. Germination data – Half-normal plot with simulation envelopes for deviance residuals of: (a) binomial model; (b) quasi-binomial model; (c) Beta Binomial model; all of them with maximal linear predictor (Equation (7)).

this is that there is more variability than the binomial model accommodates, and hence the data is overdispersed.

As suggested by Demétrio et al. (2014), as a next step we may try to accommodate the extra variability by estimating the dispersion parameter with a quasi-binomial model.

Fitting the quasi-binomial model

As we can see from Equation (4), the quasi-likelihood approach estimates the dispersion parameter to allow for greater variability. The quasi-binomial model (M2), with the same linear predictor (7), is easily fitted in R using the quasi-binomial family in the `glm` function, and the `hnp` function can still be used to assess goodness-of-fit (Box 2).

Box 2. R code to fit the quasi-binomial model to the germination data.

```
# fitting the quasi-binomial model (M2) using block and fung as
# factor, dose as numeric and the linear predictor (7)
M2 <- glm(resp ~ BLOCK + FUNG*DOSE, family=quasibinomial,
          data=germination).
# check adequacy of model using hnp
hnp(M2, paint.out=TRUE, print=TRUE)
```

Again, there is strong evidence of an inadequate model fit, with 80.71% of the observed residuals lying outside the simulated envelope (Figure 2(b)).

Fitting the beta-binomial model

The beta-binomial distribution is an alternative approach to account for overdispersion. We can easily fit the beta-binomial model (M3) with the same linear predictor (7) in R (Box 3) by making use of the `aods3` package (Lesnoff & Lancelot, 2013).

Box 3. R code to fit the beta binomial model to the germination data.

```
require(aods3)
# fitting the beta-binomial model (M3) using block and fung as
# factor, dose as numeric and the linear predictor (7)
M3 <- aodml (resp ~ BLOCK+FUNG*DOSE, family='bb', data=germination)
# check adequacy of model using hnp
hnp (M3, paint.out=TRUE, print=TRUE)
```

It appears that this model provides a better fit to the data when compared to the quasi-binomial model. However, there is still a considerable amount of points outside of the simulated envelope when looking at the half-normal plot of the deviance residuals (Figure 2(c)). The next step is to try a model that incorporates random effects to capture this extra-variation.

Fitting the logistic-normal model

In each of the blocks, the same suspensions were used for the inoculation. Then we can add a random effect to account for the correlation between proportions taken on the same isolate and block. Using the logistic-normal model (M4), we can add to the linear predict (7) a normal random effect on the intercept

$$\eta_{jik} = \gamma_j + \beta_{0i} + \beta_{1i}t_k + b_{0ij}, \quad j = 1, \dots, 4, i = 1, \dots, 14 \quad \text{and} \quad k = 1, \dots, 5, \quad (8)$$

where b_{0ij} is the random effect for the intercept of the i -th isolate and j -th block with variances σ_I^2 . We may easily fit different logistic-normal model (M5) in which a random effect to slope is also included in the linear predictor (8)

$$\eta_{jik} = \gamma_j + \beta_{0i} + \beta_{1i}t_k + b_{0ij} + b_{1ij}t_k, \quad (9)$$

where b_{1ij} is the random effect for the slope of the i -th isolate and j -th block with variances σ_S^2 and $\begin{bmatrix} b_{0ij} \\ b_{1ij} \end{bmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right]$ with $\Sigma = \begin{bmatrix} \sigma_I^2 & \sigma_{IS} \\ \sigma_{IS} & \sigma_S^2 \end{bmatrix}$. These two normal random effects will account for part of the variability induced by the use of the different fungal suspensions.

Another practical explanation for the extra-variation could be the random variation of the conidia concentration. This can be taken into account in the model by adding in the linear predictor (9) an observation-level random effect (M6)

$$\eta_{jik} = \gamma_j + \beta_{0i} + \beta_{1i}t_k + \beta_{2j} + b_{0ij} + b_{1ij}t_k + z_{jik}, \quad (10)$$

where $z_{jik} \sim N(0, \sigma_0^2)$ is the observation level random effect. For the models presented in Equations (9) and (10) we assume $\begin{bmatrix} b_{0ij} \\ b_{1ij} \end{bmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right]$ with $\Sigma = \begin{bmatrix} \sigma_I^2 & \sigma_{IS} \\ \sigma_{IS} & \sigma_S^2 \end{bmatrix}$.

We now consider fitting these models making use of the `lme4` package (Bates, Maechler, Bolker, & Walker, 2014) as showed in Box 4. When adding random effects, some convergence problems may arise due to numerical integration problems, so it is important to take into consideration different approaches. Here we are using the

bobyqa method (Powell, 2009) with maximum number of function evaluations allowed equal to 500.000, as it yielded better convergence in our case. The purpose of bobyqa is to minimise a function of many variables by a trust region method that forms quadratic models by interpolation. Another option is to use the Nelder-Mead method, which is the in `glmer`.

Box 4. R code to fit the logistic-normal models (M4) and (M5) with linear predictors shown in Equations (8) and (9).

```
require(lme4)
# fitting logistic normal-model (M4) with the linear predictor (8)
M4 <- glmer(resp ~ BLOCK+FUNG*DOSE+(1|FUNG:BLOCK), family=binomial,
            data=germination, control=glmerControl(optimizer='bobyqa',
            optCtrl=list(maxfun=500000)))

# fitting logistic normal-model (M5) with the linear predictor (9)
M5 <- glmer(resp ~ BLOCK+FUNG*DOSE+(DOSE|FUNG:BLOCK), family=binomial,
            data=germination, control=glmerControl(optimizer='bobyqa',
            optCtrl=list(maxfun=500000)))

# check adequacy of models using hnp
hnp(M4, paint.out=TRUE, print=TRUE)
hnp(M5, paint.out=TRUE, print=TRUE)
```

The inclusion of these random effects alone were not enough to explain the extra variation (see Figure 3(a) and (b)). A good alternative, with a practical explanation, to model this extra-variation is to add in the linear predictor (9) a random effect to each experimental unit (M6). This could explain variation induced by random variation of conidia concentration by dish when preparing the solutions (Box 5).

Box 5. R code to fit the logistic-normal model (M6) with the linear predictor shown in Equation (10).

```
# fitting logistic normal-model (M6) with the linear predictor (10)
ind <- gl(nrow(germination), 1)
M6 <- glmer(resp ~ BLOCK+FUNG*DOSE + (DOSE|FUNG:BLOCK) + (1|ind),
            family=binomial, data=germination,
            control=glmerControl(optimizer='bobyqa',
            optCtrl=list(maxfun=500000)))

# check adequacy of models using hnp
hnp(M6, paint.out=TRUE, print=TRUE)
```

The half-normal plot (Figure 3(c)) shows no evidence of an inadequate model with most observed residuals lying inside the simulated envelope. It is important to note that every time we produce a simulated envelope, it will slightly change and hence the number of observations lying within the envelope may differ. Now we can assess the significance of the variance components and compare fitted models with one another (Box 6). The suitability of the logistic-normal model is consistently verified by comparing the Akaike Information Criteria (AIC) (Akaike, 1973) for the binomial (M1, AIC = 6437), beta-binomial (M3, AI = 1896), and logistic-normal models (M7, AIC = 1749.3). A smaller AIC gives evidence of a better-fitted model, and hence the logistic-normal model would be selected here.

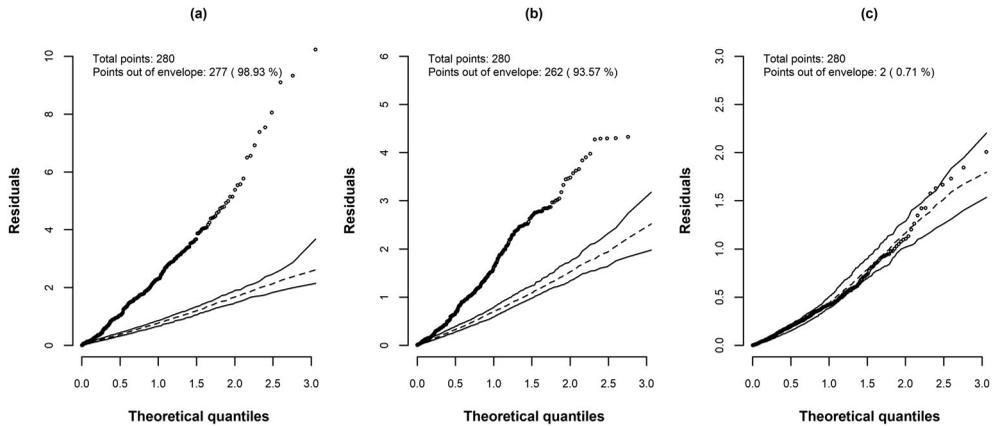


Figure 3. Germination data – Half-normal plots with simulation envelopes of deviance residuals for (a) logistic-normal model with linear predictor (Equation (8)); (b) logistic-normal model with linear predictor (Equation (9)) and (c) logistic-normal model with linear predictor (Equation (10)).

Box 6. R code for comparing models (M4), (M5) and (M6).

```
# get deviance for models
-2*logLik(M4)
'log Lik.' 3036.34
-2*logLik(M5)
'log Lik.' 2110.31
Dif1 <- 3036.34-2110.31
-2*logLik(M6)
'log Lik.' 1679.27
Dif2 <- 2110.31-1679.27

# p-value for comparing M4 and M5
p <- 0.5*pchisq(Dif1,df=1, lower.tail=F) + 0.5*pchisq(Dif1,df=2,
lower.tail=F);p
[ 1] 4.212566e-202 # M5 (most complex) is selected because p < 0.05

# p-value for comparing M5 and M6
p <- 0.5*pchisq(Dif2,df=0, lower.tail=F) + 0.5*pchisq(Dif2,df=1,
lower.tail=F);p
[ 1] 1.315649e-96 # M6 (most complex) is selected because p < 0.05
```

The linear predictor (8) includes random intercepts only, assuming all measurements on the i -th isolate and j -th block are correlated. The linear predictor (9) includes random linear dose effects b_{0ij} correlated with the random intercepts b_{1ij} and the linear predictor (10) includes all effects in the linear predictor (9) besides the observation level random effect z_{ijk} . Here, by using likelihood-ratio tests, we can conclude that model (M6) is the most suitable to analyze the data (see Table 1).

After the selection of random effects is concluded, we may now test for parallelism of linear predictors. In this experiment, if we select the linear predictor with separate lines, this will mean that each isolates curve presents different intercepts and slopes (see Box 7). If we select the parallel lines linear predictor, the curves present different intercepts but the same slope for all isolates (linear predictor (11)). Finally, by selecting the

Table 1. Summary of the results of likelihood ratio (LR) tests for the comparison of generalised linear mixed models (M4), (M5) and (M6).

Model	Distribution of random effects	Covariance matrix	Null Hypothesis	LR test ($D_q - D_p$)
M4:	$b_{0ij} \sim N(0, \Sigma)$	$\Sigma = (\sigma_I^2)$		
M5:	$\begin{bmatrix} b_{0ij} \\ b_{1ij} \end{bmatrix} \sim N_2 \left[\begin{matrix} 0 \\ 0 \end{matrix}, \Sigma \right]$	$\Sigma = \begin{pmatrix} \sigma_I^2 & \sigma_{IS} \\ \sigma_{IS} & \sigma_S^2 \end{pmatrix}$	$H_0: \sigma_S^2 = \sigma_{IS}^2 = 0$	3036.34 – 2110.3 = 926 Reference distribution: $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ $p < 0.0001$
M6:	$\begin{bmatrix} b_{0ij} \\ b_{1ij} \\ z_{jik} \end{bmatrix} \sim N \left[\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}, \Sigma \right]$	$\Sigma = \begin{pmatrix} \sigma_I^2 & \sigma_{IS} & 0 \\ \sigma_{IS} & \sigma_S^2 & 0 \\ 0 & 0 & \sigma_0^2 \end{pmatrix}$	$H_0: \sigma_0^2 = 0$	2110.3 – 1679.3 = 431 Reference distribution: $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ $p < 0.0001$

coincident line linear predictor, all isolates present the same intercept and slope (linear predictor (12)).

$$n_{jik} = \gamma_j + \beta_{0i} + \beta_1 t_k + \beta_{2j} + b_{0ij} + b_{1ij} t_k + z_{jik} \tag{11}$$

$$n_{jik} = \gamma_j + \beta_0 + \beta_1 t_k + \beta_{2j} + b_{0ij} + b_{1ij} t_k + z_{jik} \tag{12}$$

The parallel lines logistic-normal model (M7) was selected through likelihood-ratio tests (see Table 2). The fitted curves presented in Figure 4 are constructed using the linear predictor (11) and the mean of the estimated effects of the blocks, presented in Table 3 (see Box 8) The half-normal plot to model (M7) is presented in Figure 5.

Box 7. R code to fit linear predictors (11) and (12), and LR test for models (M6), (M7) and (M8).

```
# fitting the logistic normal-model (M7) with fung, dose and without interaction
effect - linear predictor (11) (Parallel lines)
M7 <- glmer(resp ~ BLOCK+FUNG+DOSE + (DOSE|FUNG:BLOCK) + (1|ind),
            family=binomial, data=germination,
            control=glmerControl(optimizer='bobyqa',
                                optCtrl=list(maxfun=50000)))
# fitting the logistic normal-model (M8) with dose, without fung and interaction
effect - linear predictor (12) (Coincident lines)
M8 <- glmer(resp ~ BLOCK+DOSE + (DOSE|FUNG:BLOCK) + (1|ind),
            family=binomial, data=germination,
            control=glmerControl(optimizer='bobyqa',
                                optCtrl=list(maxfun=50000)))

# LR tests for nested models
anova(M6, M7, M8)
```

Table 2. Likelihood-ratio tests for the logistic-normal models with separate, parallel and coincident linear predictors.

Test	χ^2	df	p-value
Separate vs parallel	12.46	13	0.4904
Parallel vs coincident	27.49	13	0.0090

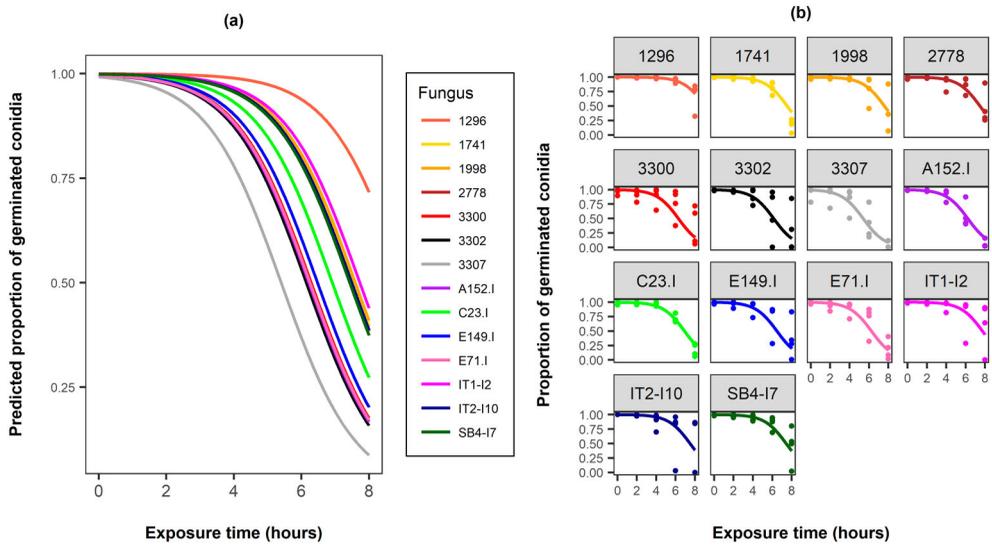


Figure 4. Fitted proportions using (M7a) for (a) each isolate in the same plot and (b) each isolate in different plots with points representing the observed values.

Table 3. Parameter estimates and Standard Errors for the selected model (M7), without intercept (M7a), before and after grouping the treatments (M12a).

Model M7a – Before Grouping		Model M12a – After Grouping	
Effect	Estimate (Std. Error)	Effect	Estimate (Std. Error)
Block 1	8.2183(0.6456)	Block 1	8.2307(0.6646)
Block 2	7.3519(0.6428)	Block 2	7.3440(0.6606)
Block 3	8.3111(0.6771)	Block 3	8.2635(0.6906)
Block 4	8.6385(0.6510)	Block 4	8.6476(0.6697)
Fung 1741	-1.3479(0.7608)	Group 2	8.6473(0.6697)
Fung 1998	-1.2948(0.7455)	Group 3	-2.6047(0.6115)
Fung 2778	-1.4004(0.7457)	Exposure time	-0.9002(0.0542)
Fung 3300	-2.4699(0.7361)		
Fung 3302	-2.4699(0.7361)		
Fung 3307	-3.2694(0.7398)		
Fung A152-I	-2.5329(0.7397)		
Fung C23-I	-2.5329(0.7397)		
Fung 149-I	-2.3015(0.7302)		
Fung E71-I	-2.5032(0.7298)		
Fung IT1-I2	-1.1730(0.7507)		
Fung IT2-I10	-1.3875(0.7767)		
Fung SB4-I7	-1.4418(0.7440)		
Exposure time	-0.9003(0.0541)		
Observation (σ_0^2)	0.9024	Observation (σ_0^2)	0.8958
Intercept Block:Fung (σ_7^2)	1.7906	Intercept Block:Fung (σ_7^2)	1.7272
Slope Block:Fung (σ_3^2)	0.1078	Slope Block:Fung (σ_3^2)	0.1092
Covariance (σ_{75})	-0.3602	Covariance (σ_{75})	-0.3560

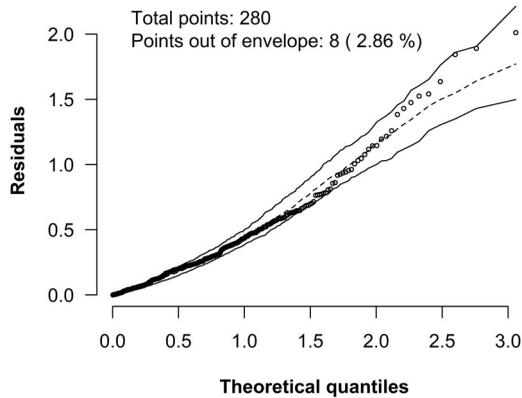


Figure 5. Germination data – Half-normal plot with simulation envelopes of deviance residuals for logistic-normal model with the parallel lines (M7) with linear predictor (Equation (11)).

Box 8. R code to create the function to calculate exposure time for germination of 75% and 95% germination of conidia.

```
# fitting logistic normal-model without intercept (-1) to obtain every
# block effects
M7a <- glmer (resp ~ -1+BLOCK+FUNG+DOSE + (DOSE|FUNG:BLOCK) + (1|ind),
             family=binomial, data= germination,
             control=glmerControl(optimizer='bobyqa',
                                 optCtrl=list(maxfun=50000)))

# get block effects 1 to 4, fung effects 2 to 14 and Dose.
block<- (summary(M7a)$coef[ 1:4,1] )
fungus<- (summary(M7a)$coef[ 5:17,1] )
slope<- (summary(M7a)$coef[ 18,1] )

# the next functions presented only works for this specific model!

# creating function to calculate the estimated exposure time from
# 1° Fung Isaria 1296
dg_1296 = function(percentil){
  dg_1296 = as.data.frame(round((log(percentil/(1-percentil)) -
    (1/4)*(block[ 1]+block[ 2]+block[ 3]+block[ 4] ))/slope, 2))
  colnames(dg_1296) = paste('germination',percentil*100,'%')
  return (dg_1296)
}
cbind(dg_1296(0.75),dg_1296(0.9))

# creating function to calculate the estimated exposure time from other # fungus
dg = function(percentil){
  dg = as.data.frame(round((log(percentil/(1-percentil)) -
    (1/4)*(block[ 1]+block[ 2]+block[ 3]+block[ 4] )-fungus[ 1:13] ))/slope, 2))
  colnames(dg) = paste('germination',percentil*100,'%')
  return (dg)
}
cbind(dg(0.75),dg(0.9))
```

We may now use the selected model to determine the exposure time that resulted in some specified proportion of germination, say 75% and 90%. We can obtain these estimates using the code presented in [Box 8](#).

Table 4. Exposure time (in hours) for 75 and 90% germination by fungus.

Fungus	75% germination	90% germination
	Exposure time	Exposure time
1296	7.81	6.59
IT1-I2	6.51	5.29
1998	6.37	5.15
1741	6.31	5.09
2778	6.25	5.03
IT2-I10	6.27	5.05
SB4-I7	6.21	4.99
C23.I	5.69	4.47
E149.I	5.25	4.03
3300	5.07	3.85
E71.I	5.03	3.81
A152.I	5.00	3.78
3302	4.93	3.70
3307	4.18	2.97

The *I. fumosorosea* 1296 was the most tolerant isolate because presented a time to reduce viability of 7.81 and 6.59 hours compared to the most susceptible isolate, *I. fumosorosea* 3307 which presented 4.18 and 2.97 h (see Table 4).

Grouping

Because of the similarity observed between isolates in Figure 4(a), some of them were grouped and a likelihood-ratio test was performed to test these differences. The groups were constructed according to the predicted curves in order to group the ones that had a similar behaviour. For example the test of groups to be created consisted of 5 groups (Table 5).

This test was not significant (see Box 9), hence there is no evidence against a simpler adequate model, and the isolates within groups are concluded to not differ statistically.

Box 9. R code to create and to test groupings.

```
# creating the Factor Grouping 1
germination$Grouping1 <- germination$FUNG
levels(germination$Grouping1)
[ 1] "1296" "1741" "1998" "2778" "3300" "3302"
[ 7] "3307" "A152.I" "C23.I" "E149.I" "E71.I"
[12] "IT1-I2" "IT2-I10" "SB4-I7"
levels(germination$Grouping1) <- c(1,2,2,2,4,4,5,4,3,4,4,2,2,2)
# model fitted with factor Grouping 1 instead of Fung
M8 <- glmer(resp ~ BLOCK + Grouping1 + DOSE + (DOSE|FUNG:BLOCK)
            + (1|ind), family=binomial, data=germination,
            control=glmerControl(optimizer='bobyqa',
                                optCtrl=list(maxfun=50000)))
# testing grouping (equality between isolates of the same group)
anova(M8,M7,test='Chisq')
Data: germination
Models:
M8: resp ~ BLOCK + Grouping1 + DOSE + (DOSE|FUNG:BLOCK) + (1|ind)
M7: resp ~ BLOCK + FUNG + DOSE + (DOSE|FUNG:BLOCK) + (1|ind)
   Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
M8: 13 1718.1 1765.4 -846.06   1692.1
M7: 22 1735.7 1815.7 -845.86   1691.7 0.396  9  1
# M8 (simplest) is selected, because p>0.05, i.e. the isolates within
# the groups do not differ statistically
```

Table 5. Grouping 1 – first grouping according to the proximities between the curves presented in Figure 4(a).

Group1	Group2	Group3	Group4	Group5
1296	1741	C23.I	3300	3307
	1998		A152.I	
	2778		E149.I	
	IT1-12		E71.I	
	IT2-10		3302	
	SB4-I7			

Once the five groups described above were constructed, more similarities were sought between these isolates by testing the clustering of these groups in the attempt to identify the most resistant group. Grouping 2 was created to test resistance similarities between *Isaria* 1296 and isolates of Group 2 (Table 5), but this hypothesis was rejected (see Box 10). Grouping 3 was created to test similarities between *Isaria* C23.I and Group 2 (Table 5). Grouping 4 was created to test similarities between Groups 2, 3 and 4 (Table 5), but this hypothesis was rejected (Box 10). Finally, Grouping 5 was created to test equality of Groups 4 and Group 5 (Table 5).

Box 10. R code to created and test different groupings of isolates.

```
# creating factor Grouping 2
germination$Grouping2 <- germination$FUNG
levels(germination$Grouping2) <- c(2,2,2,2,4,4,5,4,3,4,4,2,2,2)
# creating factor Grouping 3
germination$Grouping3<- germination$FUNG
levels(germination$Grouping3) <- c(1,2,2,2,4,4,5,4,2,4,4,2,2,2)
# creating factor Grouping 4
germination$Grouping4<- germination$FUNG
levels(germination$Grouping4) <- c(1,2,2,2,2,2,5,2,2,2,2,2,2,2)
# creating factor Grouping 5
germination$Grouping5<- germination$FUNG
levels(germination$Grouping5) <- c(1,2,2,2,4,4,4,4,2,4,4,2,2,2)
# model fitted with factor Grouping 2
M9 <- glmer(resp ~ BLOCK + Grouping2 + DOSE + (DOSE|FUNG:BLOCK)
+ (1|ind), family=binomial, data=germination,
control=glmerControl(optimizer='bobyqa',
optCtrl=list(maxfun=50000)))
# model fitted with factor Grouping 3
M10 <- glmer(resp ~ BLOCK + Grouping3 + DOSE + (DOSE|FUNG:BLOCK)
+ (1|ind), family=binomial, data=germination,
control=glmerControl(optimizer='bobyqa',
optCtrl=list(maxfun=50000)))
# model fitted with factor Grouping 4
M11 = glmer(resp ~ BLOCK + Grouping4 + DOSE + (DOSE|FUNG:BLOCK) + (1|ind),
family=binomial, data=germination,
control=glmerControl(optimizer='bobyqa',
optCtrl=list(maxfun=50000)))
# model fitted with factor Grouping 5
M12 = glmer(resp ~ BLOCK + Grouping5 + DOSE + (DOSE|FUNG:BLOCK) + (1|ind),
family=binomial, data=germination,
control=glmerControl(optimizer='bobyqa',
optCtrl=list(maxfun=50000)))

# testing equality between isolates of Group 2 and Isolate 1296
anova(M9,M8,test='Chisq')
Df AIC BIC logLik deviance Chisq Chi Df Pr (>Chisq)
M9: 12 1721.1 1764.7 -848.56 1697.1
```

```

M8: 13 1718.1 1765.4 -846.06 1692.1 5.0031 1 0.0253 *
# M8 (most complex) was selected, because p < 0.05, i.e.
Isolate 1296 differs from isolates in Group 2

# testing equality between isolates from Group 2 and Isolate C23.I
anova(M10,M8,test='Chisq')
  Df  AIC  BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
M10: 12 1717.2 1760.8 -846.61 1693.2
M8: 13 1718.1 1765.4 -846.06 1692.1 11.1056 1 0.293
# M10 (simplest) was selected, because p > 0.05, i.e. Isolate C23.I
does not differs from isolates in Group 2

# testing equality between Group 2 plus Isolate C23.I and Group 4
anova(M11,M10,test='Chisq')
  Df  AIC  BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
M11: 11 1727.3 1767.2 -852.63 1705.3
M10: 12 1717.2 1760.8 -846.61 1693.2 12.031 1 0.0005231 ***
#M10 (most complex) was selected, because p < 0.05, i.e. the isolates in Group 2 plus
Isolate C23.I differ from the isolates in Group 4

# testing equality between Group 4 and Isolate 3307
anova(M12,M10,test='Chisq')
  Df  AIC  BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
M12: 11 1717.3 1757.3 -847.64 1695.3
M10: 12 1717.2 1760.8 -846.61 1693.2 2.0582 1 0.1514
#M12 (simplest) was selected, because p > 0.05, i.e. the isolates in Group 4 does not
differ from the Isolate 3307

# fitting M12 without intercept (-1) to obtain every block effects
M12a = glmer(resp ~ -1+BLOCK + Grouping5 + DOSE + (DOSE|FUNG:BLOCK)
+ (1|ind), family=binomial, data=germination,
control=glmerControl(optimizer='bobyqa',
optCtrl=list(maxfun=50000)))

```

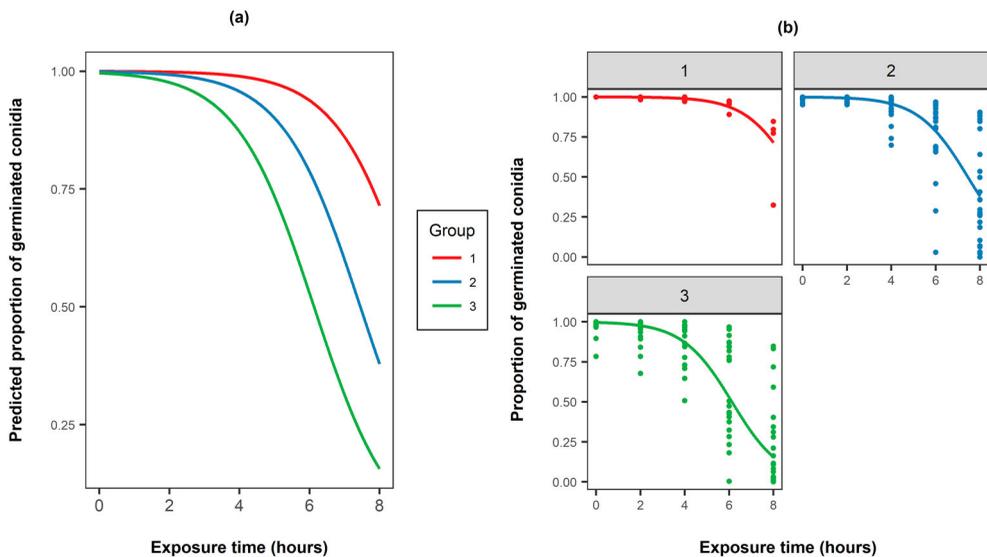


Figure 6. Fitted group proportions using a parallel lines logistic-normal for (a) each isolate in the same plot and (b) each isolate in different plots with points representing the observed values.

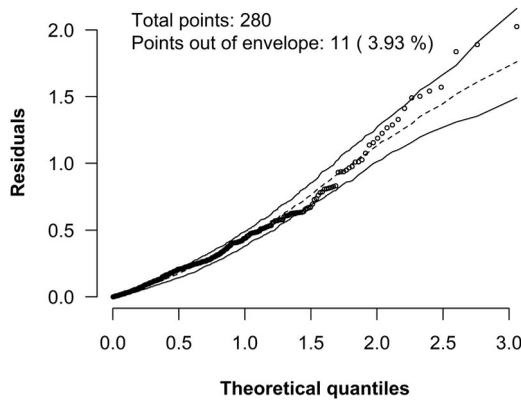


Figure 7. Germination data – Half-normal plot with simulation envelopes of deviance residuals for fitted group proportions using a parallel lines logistic-normal model (M12).

According to tests shown above, we can group the isolates in three groups in which the fungi belonging to distinct groups are significantly different at a significance level of 5%.

Group 1: *I. fumosorosea* 1296

Group 2: *I. fumosorosea* 1741, 1998, 2778, IT1-12, IT2-I10, SB4-I7, C23.I

Group 3: *I. fumosorosea* 3300, 3302, A152.I, E149.I, E71.I, 3307

Fitted proportions (M12a) are shown in Figure 6 and the half-normal plot for the deviance residuals in Figure 7. This graphics indicates that this is also an adequate model fit, and hence this model can be used to make inference on the data. The parameter estimates are given in Table 3.

Discussion

The aim of this paper was to propose a tutorial, using software R, on how to fit and to assess goodness-of-fit for a range of different models for overdispersion data. We described and showed, step by step, how to fit and interpret the simplest, binomial model, as well as the more complex logistic-normal model.

We use, as a motivation for this tutorial, a dataset from an experiment involving germination of conidia from different species of entomopathogenic fungi, focusing on how to identify and account for specific issues arising from experimental design (such as correlated observations) and extra-variation (i.e. overdispersion).

For the case study presented here, we concluded that the logistic-normal model provided a good fit to the data, and included three random effects: a random intercept per block, a random slope per block and an observation-level random effect to capture overdispersion. Due to the inclusion of three random effects, some convergence problems occurred because of numerical integration problems, so it is important to take into consideration different numerical integration approaches (e.g. Laplace approximation, Adaptive Gauss-Hermite quadrature, Monte-Carlo methods) available to tackle this problem. As a practical outcome, the selected model allowed us to identify the strain of

I. fumosorosea that presented the highest UV-B radiation tolerance. We were also able to group similar isolates using likelihood-ratio tests, which yields important information in the context of biological control.

The simpler binomial model could have led to mistaken conclusions, such as the selection of the separate linear predictors model or different groupings of isolates. The use of a GLMM solved this issue and allowed us to model the correlation between observations within each block. Hence, we advocate for the use of half-normal plots with simulation envelopes, and other goodness-of-fit assessment techniques, to ensure inference made using the fitted model is reliable.

Finally, we would like to stress that this paper does not present a recipe for analyzing this type of data, since each experiment is different and is carried out to test different types of hypotheses. It is of utmost importance that the modelling takes into account the experimental design and the nature of the data, and that the appropriate hypothesis tests are performed. Every dataset presents its own peculiarities which may guide the statistical analyses one way or another.

Note

1. If the random variables in a finite or infinite sequence X_1, X_2, \dots, X_n are independent and identically distributed, and if each random variable X_i has the Bernoulli distribution with parameter p , then it is said that X_1, X_2, \dots, X_n are *Bernoulli trials with parameter p* (Degroot & Schervish, 2012).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq - grant number 308464/2017-6), the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior – CAPES) and FAPESP (grant number 2014/03310-3).

ORCID

Máira Blumer Fatoretto  <http://orcid.org/0000-0002-2508-1586>

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Caski (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Atkinson, A. C. (1985). *Plots, transformations, and regression: An introduction to graphical methods of diagnostic regression analysis*. Oxford: Clarendon.

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, 1(7), 1–23.
- Degroot, M. H., & Schervish, M. J. (2012). *Probability and statistics*. New York: Pearson Education.
- Demétrio, C. G. B., Hinde, J., & Moral, R. A. (2014). Models for overdispersed data. In C. Ferreira, & W. Godoy (Eds.), *Ecological modelling applied to entomology. Entomology in focus* (pp. 219–259). New Delhi: Springer International Publishing.
- Faria, M., & Wraight, S. P. (2007). Mycoinsecticides and mycoacaricides: A comprehensive list with worldwide coverage and international classification of formulation types. *Biological Control*, 43, 237–256.
- Hinde, J., & Demétrio, C. G. B. (1998). Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*, 27, 151–170.
- Lesnoff, M., & Lancelot, R. (2013). analysis of overdispersed data using S3 methods. Aods3 package version 0.4-1. Retrieved from <http://CRAN.R-project.org/package=aods3>
- Moral, R. A., Hinde, J., & Demétrio, C. G. (2017). Half-Normal plots and overdispersed models in R: The hnp package. *Journal of Statistical Software*, 81(10), 1–23.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). *Generalized linear models*. London: Wiley Online Library.
- Oliveira, D. G. P., Pauli, G., Mascarin, G. M., & Delalibera, I. (2015). A protocol for determination of conidial viability of the fungal entomopathogens *Beauveria bassiana* and *Metarhizium anisopliae* from commercial products. *Journal of Microbiological Methods*, 119, 44–52.
- Paul, N. D., Rasanayagam, S., Moody, S. A., Hatcher, P. E., & Ayres, P. G. (1997). The role of interactions between trophic levels in determining the effects of UV-B on terrestrial ecosystems. *Plant Ecology*, 128, 296–308.
- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. (Report No. DAMTP 2009/NA06). England, Department of Applied Mathematics and Theoretical Physics, University of Cambridge.
- R. Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Verbeke, G., & Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, 59(2), 254–262.
- Zimmermann, G. (2008). The entomopathogenic fungi *Isaria farinosa* (formerly *Paecilomyces farinosus*) and the *Isaria fumosorosea* species complex (formerly *Paecilomyces fumosoroseus*): biology, ecology and use in biological control. *Biocontrol Science and Technology*, 18(9), 865–901.