The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded

Shinichi Nakagawa[1,2], Paul C. D. Johnson[3] and Holger Schielzeth[4]

[1] Evolution & Ecology Research Centre, and School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[2] Diabetes and Metabolism Division, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia

[3] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Graham Kerr Building, Glasgow G12 8QQ, UK

[4] Population Ecology Group, Institute of Ecology, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany

Running head: R^2 and ICC from GLMMs

Address correspondence to S. Nakagawa. Email: s.nakagawa@unsw.edu.au; H. Schielzeth. Email: holger.schielzeth@uni-jena.de

1

1 **Abstract**

2 The coefficient of determination R^2 quantifies the proportion of variance explained by a statistical

3 model and is an important summary statistic of biological interest. However, estimating R^2 for

4 generalized linear mixed models (GLMMs) remains challenging. We have previously introduced a

5 version of R^2 that we called R^2_{GLMM} for Poisson and binomial GLMMs, but not for other

6 distributional families. Similarly, we earlier discussed how to estimate intra-class correlation

7 coefficients ICC using Poisson and binomial GLMMs. In this article, we expand our methods to all

8 other non-Gaussian distributions, in particular to negative binomial and gamma distributions that

9 are commonly used for modelling biological data. While expanding our approach, we highlight two

10 useful concepts for biologists, Jensen's inequality and the delta method, both of which help us in

11 understanding the properties of GLMMs. Jensen's inequality has important implications for

12 biologically meaningful interpretation of GLMMs, while the delta method allows a general

13 derivation of variance associated with non-Gaussian distributions. We also discuss some special

14 considerations for binomial GLMMs with binary or proportion data. We illustrate the

15 implementation of our extension by worked examples from the field of ecology and evolution in the

16 R environment. However, our method can be used across disciplines and regardless of statistical

17 environments.

18 ***Key words:*** *repeatability, heritability, goodness of fit, model fit, variance decomposition, reliability*

19 *analysis.*

20

2

21

1. Introduction

23 One of the main purposes of linear modelling is to understand the sources of variation in biological

24 data. In this context, it is not surprising that the coefficient of determination R^2 is a commonly

25 reported statistic because it represents the proportion of variance explained by a linear model. The

26 intra-class correlation coefficient ICC is a related statistic that quantifies the proportion of variance

27 explained by a grouping (random) factor in multilevel/hierarchical data. In the field of ecology and

28 evolution, a type of ICC is often referred to as repeatability R, where the grouping factor is often

29 individuals that have been phenotyped repeatedly [1, 2]. We have reviewed methods for estimating

30 R^2 and ICC in the past, with a particular focus on non-Gaussian response variables in the context of

31 biological data [2, 3]. These previous articles featured generalized linear mixed-effects models

32 (GLMMs) as the most versatile engine for estimating R^2 and ICC (specifically R^2_{GLMM} and

33 ICC_{GLMM}). Our descriptions were limited to random-intercept GLMMs, but Johnson [4] has

34 recently extended the methods to random-slope GLMMs, widening the applicability of these

35 statistics (see also, [5, 6]).

36 However, at least one important issue seems to remain. Currently these two statistics are only

37 described for binomial and Poisson GLMMs. Although these two types of GLMMs are arguably the

38 most popular [7], there are other families of distributions that are commonly used in biology, such

39 as negative binomial and gamma distributions [8, 9]. In this article, we revisit and extend R^2_{GLMM}

40 and ICC_{GLMM} to more distributional families with a particular focus on negative binomial and

41 gamma distributions. In this context, we discuss Jensen's inequality and two variants of the delta

42 method, which are hardly known among biologists. These concepts are useful not only for

43 generalizing our previous methods, but also for interpreting the results of GLMMs. Furthermore,

44 we refer to some special considerations when obtaining R^2_{GLMM} and ICC_{GLMM} from binomial

45 GLMMs for binary and proportion data, which we did not discuss in the past [2, 3]. We provide

46 worked examples inspired from the field of ecology and evolution, focusing on implementation in

47   the R environment [10] and finish by referring to two alternative approaches for obtaining R^2 and

48   ICC from GLMMs along with a cautionary note.

2. Definitions of R^2_{GLMM}, ICC$_{GLMM}$ and overdispersion

50   To start with, we present R^2_{GLMM} and ICC$_{GLMM}$ for a simple case of Gaussian error distributions

51   based on a linear mixed-effects model (LMM, hence also referred to as R^2_{LMM} and ICC$_{LMM}$).

52   Imagine a two-level dataset where the first level corresponds to observations and the second level to

53   some grouping/clustering factor (e.g. individuals with repeated measurements) with k fixed effect

54   covariates. The model can be written as (referred to as Model 1):

$$y_{ij} = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + \varepsilon_{ij}, \tag{2.1}$$

$$\alpha_i \sim \text{Gaussian}(0, \ \sigma_\alpha^2), \tag{2.2}$$

$$\varepsilon_{ij} \sim \text{Gaussian}(0, \ \sigma_\varepsilon^2), \tag{2.3}$$

58   where y_{ij} is the jth observation of the ith individual, x_{hij} is the jth value of the ith individual for the

59   hth of k fixed effects predictors, β_0 is the (grand) intercept, β_h is the regression coefficient for the

60   hth predictor, α_i is an individual-specific effect, assumed to be normally distributed in the

61   population with the mean and variance of 0 and σ_α^2, ε_{ij} is an observation-specific residual, assumed

62   to be normally distributed in the population with mean and variance of 0 and σ_ε^2, respectively. For

63   this model, we can define two types of R^2 as:

$$R^2_{LMM(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{2.4}$$

$$R^2_{LMM(c)} = \frac{\sigma_f^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{2.5}$$

$$\sigma_f^2 = \text{var}\left(\sum_h^k \beta_h x_{hij}\right), \tag{2.6}$$

67   where $R^2_{\text{LMM}(m)}$ represents the marginal R^2, which is the proportion of the total variance explained by

68   the fixed effects, $R^2_{\text{LMM}(c)}$ represents the conditional R^2, which is the proportion of the variance

69   explained by both fixed and random effects, and σ^2_f is the variance explained by fixed effects [11].

70   Since marginal and conditional R^2 differ only in whether the random effect variance is included in

71   the numerator, we avoid redundancy and present equations only for marginal R^2 in the following.

72   Similarly, there are two types of ICC:

73    $$\text{ICC}_{\text{LMM(adj)}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\varepsilon} \qquad\qquad (2.7)$$

74    $$\text{ICC}_{\text{LMM}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_f + \sigma^2_\varepsilon} \qquad\qquad (2.8)$$

75   If no fixed effects are fitted (other than the intercept), $\sigma^2_f = 0$ so that $\text{ICC}_{\text{LMM(adj)}}$ equals ICC_{LMM}. In

76   such a case, the ICC should not be called 'adjusted'(*sensu* [2]). For an ICC value to be adjusted for

77   a source of variance, that variance must be more than 0 and omitted from the ICC calculation. Since

78   the two versions of ICC differ only in whether the fixed effect variance, calculated as in equation

79   (2.6), is included in the denominator, we avoid redundancy and present equations only for adjusted

80   ICC in the following.

81   One of the main difficulties in extending R^2 from LMMs to GLMMs is defining the residual

82   variance σ^2_ε. For binomial and Poisson GLMMs with an additive dispersion term, we have

83   previously stated that σ^2_ε is equivalent to $\sigma^2_e + \sigma^2_d$ where σ^2_e is the variance for the additive

84   overdispersion term, and σ^2_d is the distribution-specific variance [2, 3]. Here, overdispersion

85   represents the excess variation relative to what is expected from a certain distribution and can be

86   estimated by fitting an observation-level random effect (OLRE; see, [12, 13]). Alternatively,

87   overdispersion in GLMMs can be implemented using a multiplicative overdispersion term [14]. In

88   such an implementation, we stated that σ^2_ε is equivalent to $\omega \cdot \sigma^2_d$ where ω is a multiplicative

89   dispersion parameter estimated from the model [2]. However, obtaining σ_d^2 for specific

90   distributions is not always possible, because in many families of GLMMs, σ_ε^2 (observation-level

91   variance) cannot be clearly separated into σ_e^2 (overdispersion variance) and σ_d^2 (distribution-

92   specific variance). It turns out that binomial and Poisson distributions are special cases where σ_d^2

93   can be usefully calculated, because either all overdispersion is modelled by an OLRE (additive

94   overdispersion) or by a single multiplicative overdispersion parameter (multiplicative

95   overdispersion). This is not the case for other families. However, as we will show below, we can

96   always obtain the GLMM version of σ_ε^2 (on the latent scale) directly. We refer to this generalised

97   version of σ_ε^2 as 'the observation-level variance' here rather than the residual variance (but we keep

98   the notation σ_ε^2). Note that the observation-level variance, σ_ε^2, should not be confused with the

99   variance associated with OLRE, which estimates σ_e^2 and can be considered to be a part of σ_ε^2.

3. Extension of R^2_{GLMM} and ICC$_{\text{GLMM}}$

101   We now define R^2_{GLMM} and ICC$_{\text{GLMM}}$ for a quasi-Poisson (may also be referred to as overdispersed-

102   Poisson) GLMM, because the quasi-Poisson distribution is an extension of Poisson distribution [15,

103   16], and is similar to the negative binomial distribution at least in their common applications [9, 17].

104   Imagine count data repeatedly measured from a number of individuals with associated data on k

105   covariates. We fit a quasi-Poisson (QP) GLMM with the log link function (Model 2):

106     $y_{ij} \sim \text{QP}(\lambda_{ij},\ \omega)$,   (3.1)

107     $\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k}\beta_h x_{hij} + \alpha_i$,   (3.2)

108     $\alpha_i \sim \text{Gaussian}(0,\ \sigma_\alpha^2)$,   (3.3)

109   where y_{ij} is the jth observation of the ith individual and y_{ij} follows a quasi-Poisson distribution with

110   two parameters, λ_{ij} and ω [15, 16], $\ln(\lambda_{ij})$ is the latent value for the jth observation of the ith

111   individual, ω is the overdispersion parameter (when the multiplicative dispersion parameter ω is 1,

112   the model becomes a standard Poisson GLMM), α_i is an individual-specific effect, assumed to be

113   normally distributed in the population with the mean and variance of 0 and σ_α^2, respectively (as in

114   Model 1), and the other symbols are the same as above. Quasi-Poisson distributions have a mean of

115   λ and a variance of $\lambda\omega$ (Table 1). For such a model, we can define $R^2_{\text{GLMM}(m)}$ and (adjusted)

116   ICC_{GLMM} as:

117
$$R^2_{\text{QP-ln}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1 + \omega/\lambda)}, \tag{3.4}$$

118
$$\text{ICC}_{\text{QP-ln}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1 + \omega/\lambda)}, \tag{3.5}$$

119   where the subscript of R^2 and ICC denote the distributional family, here QP-ln for quasi-Poisson

120   distribution with log link, the term $\ln(1 + \omega/\lambda)$ corresponds to the observation-level variance σ_ε^2

121   (Table 1, for derivation see Appendix S1), ω is the overdispersion parameter, and λ is the mean

122   value of λ_{ij}. We discuss how to obtain λ below (Section 5).

123   The calculation is very similar for a negative binomial (NB) GLMM with the log link (Model 3):

124    $$y_{ij} \sim \text{NB}(\lambda_{ij}, \ \theta), \tag{3.6}$$

125    $$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i, \tag{3.7}$$

126    $$\alpha_i \sim \text{Gaussian}(0, \ \sigma_\alpha^2), \tag{3.8}$$

127   where y_{ij} is the jth observation of the ith individual and y_{ij} follows a negative binomial distribution

128   with two parameters, λ_{ij} and θ, where θ is the shape parameter of the negative binomial distribution

129   (given by the software often as the dispersion parameter), and the other symbols are the same as

130   above. The parameter θ is sometimes referred to as 'size'. Negative binomial distributions have a

131   mean of λ and a variance of $\lambda + \lambda^2/\theta$ (Table 1). $R^2_{\text{GLMM}(m)}$ and (adjusted) ICC_{GLMM} for this model

132   can be calculated as:

7

133 $$R^2_{\text{NB}-\ln(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1 + 1/\lambda + 1/\theta)},$$ (3.9)

134 $$\text{ICC}_{\text{NB}-\ln} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1 + 1/\lambda + 1/\theta)},$$ (3.10)

135 Finally, for a gamma GLMM with the log link (Model 4):

136 $$y_{ij} \sim \text{gamma}(\lambda_{ij}, \nu),$$ (3.11)

137 $$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i,$$ (3.12)

138 $$\alpha_i \sim \text{Gaussian}(0, \sigma_\alpha^2),$$ (3.13)

139 where y_{ij} is the jth observation of the ith individual and y_{ij} follows a gamma distribution with two

140 parameters, λ_{ij} and ν, where ν is the shape parameter of the gamma distribution (sometimes

141 statistical programs report $1/\nu$ instead of ν; also note that the gamma distribution can be

142 parameterized in alternative ways, Table 1). Gamma distributions have a mean of λ and a variance

143 of λ^2/ν (Table 1). $R^2_{\text{GLMM}(m)}$ and (adjusted) ICC_{GLMM} can be calculated as:

144 $$R^2_{\text{gamma}-\ln(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1 + 1/\nu)},$$ (3.15)

145 $$\text{ICC}_{\text{gamma}-\ln} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1 + 1/\nu)},$$ (3.16)

146 4. Obtaining the observation-level variance by the 'first' delta method

147 For overdispersed Poisson, negative binomial and gamma GLMMs with log link, the observation-

148 level variance σ_ε^2 can be obtained via the variance of the log-normal distribution (Appendix S1).

149 This is the approach that has led to the terms presented above. There are two more alternative

150 methods to obtain the same target: the delta method and the trigamma function. The two

151 alternatives have different advantages and we will therefore discuss them in some detail in the

152 following.

8

153   The delta method for variance approximation uses a first order Taylor series expansion, which is

154   often employed to approximate the standard error (error variance) for transformations (or functions)

155   of a variable x when the (error) variance of x itself is known (see [18]; for an accessible reference

156   for biologists, [19]). The delta method for variance approximation can be written as:

157    $$\mathrm{var}[f(x)] \approx \mathrm{var}[x]\left(\frac{d}{dx}f(x)\right)^2,$$ (4.1)

158   where x is a random variable (typically represented by observations), f represents a function (e.g.

159   log or square-root), var denotes variance, and d/dx is a (first) derivative with respect to variable x.

160   Taking derivatives of any function can be easily done using the R environment (examples can be

161   found in the Appendices). It is the delta method that Foulley and colleagues [20] used to derive the

162   distribution-specific variance σ_d^2 for Poisson GLMMs as $1/\lambda$ (see also [21]). Given that $\mathrm{var}[y] = \lambda$

163   in the case of Poisson distributions and $d\ln(\lambda)/dx = 1/\lambda$, it follows that

164    $\mathrm{var}[\ln(y)] \approx \lambda(1/\lambda)^2 = 1/\lambda$ (note that for Poisson distributions without overdispersion, σ_d^2 is equal

165   to σ_ε^2 because $\sigma_e^2 = 0$).

166   One clear advantage of the delta method is its flexibility. We can easily obtain the observation-level

167   variance σ_ε^2 for all kinds of distributions/link functions. For example, by using the delta method, it

168   is straightforward to obtain σ_ε^2 for the Tweedie distribution, which has been used to model non-

169   negative real numbers in ecology (e.g., [22, 23]). For the Tweedie distribution, the variance on the

170   observed scale has the relationship $\mathrm{var}[y] = \varphi\mu^p$ where μ is the mean on the observed scale and φ is

171   the dispersion parameter, comparable to λ and ω in equation (3.1), and p is a positive constant

172   called an index parameter. Therefore, when used with the log-link function, σ_ε^2 can be

173   approximated by $\varphi\mu^{(p-2)}$ according to equation (4.1). The log-normal approximation $\ln(1+\varphi\mu^{(p-2)})$

174   is also possible (see Appendix S1; Table 1).

9

175   The use of the trigamma function ψ_1 is limited to distributions with log link, but it is considered to

176   provide the most accurate estimate of the observation level variance σ_ε^2 in those cases. This is

177   because the variance of a gamma-distributed variable on the log scale is equal to $\psi_1(\nu)$ where ν is

178   the shape parameter of the gamma distribution [24] and hence σ_ε^2 is $\psi_1(\nu)$. At the level of the

179   statistical parameters (Table 1; on the 'expected data' scale; *sensu* [25]; see their Figure 1), Poisson

180   and negative binomial distributions can both be seen as special cases of gamma distributions, and

181   σ_ε^2 can be obtained using the trigamma function (Table 1). For example, σ_ε^2 for the Poisson

182   distribution is $\psi_1(\lambda)$ with the speciality that in the case of Poisson distributions $\sigma_\varepsilon^2 = \sigma_d^2$. As we

183   show in Appendix S2, $\ln(1+1/\lambda)$ (log-normal approximation), $1/\lambda$ (delta method approximation) and

184    $\psi_1(\lambda)$ (trigamma function) give similar results when λ is greater than 2. Our recommendation is to

185   use the trigamma function for obtaining σ_ε^2 whenever this is possible.

186   The trigamma function has been previously used to obtain observation-level variance in

187   calculations of heritability (which can be seen as a type of ICC although in a strict sense, it is not;

188   see [25]) using negative binomial GLMMs ([24, 26]; cf. [25]). Table 1 summarises observation-

189   level variance σ_ε^2 for overdispersed Poisson, negative binomial and gamma distributions for

190   commonly used link functions.

191   ## 5. How to estimate λ from data

192   For some calculations, we require an estimate of the global expected value λ. Imagine a Poisson

193   GLMM with log link and additive overdispersion fitted as an observation-level random effect

194   (Model 5):

195    $y_{ij} \sim \text{Poisson}(\lambda_{ij}),$ (5.1)

196    $\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{p} \beta_h x_{hij} + \alpha_i + e_{ij},$ (5.2)

10

197     $\alpha_i \sim \text{Gaussian}(0, \ \sigma_\alpha^2)$,   (5.3)

198     $e_{ij} \sim \text{Gaussian}(0, \ \sigma_e^2)$,   (5.4)

199   where y_{ij} is the jth observation of the ith individual, and follows a Poisson distribution with the

200   parameter λ_{ij}, e_{ij} is an additive overdispersion term for jth observation of the ith individual, and the

201   other symbols are the same as above. Poisson distributions have a mean of λ and a variance of λ (cf.

202   Table 1). Using the log-normal approximation $R^2_{\text{GLMM}(m)}$ and (adjusted) ICC_{GLMM} can be calculated

203   as:

204     $R^2_{\text{P-ln}(m)} = \dfrac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_e^2 + \ln(1 + 1/\lambda)}$,   (5.5)

205     $\text{ICC}_{\text{P-ln}} = \dfrac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_e^2 + \ln(1 + 1/\lambda)}$,   (5.6)

206   where, as mentioned above, the term $\ln(1 + 1/\lambda)$ is σ_ε^2 (or σ_d^2) for Poisson distributions with the log

207   link (Table 1).

208   In our earlier papers, we proposed to use the exponential of the intercept, $\exp(\beta_0)$ (from the

209   intercept-only model) as an estimator of λ [2, 3]; note that $\exp(\beta_0)$ from models with any fixed

210   effects will often be different from $\exp(\beta_0)$ from the intercept-only model. We also suggested that it

211   is possible to use the mean of observed values y_{ij}. Unfortunately, these two recommendations are

212   often inconsistent with each other. This is because, given Model 5 (and all the models in the

213   previous section), the following relationships hold:

214     $\exp(\beta_0) \le \text{E}[y_{ij}]$,   (5.7)

215     $\text{E}[\lambda_{ij}] = \exp(\beta_0 + 0.5\sigma_\tau^2)$,   (5.8)

216     $\text{E}[y_{ij}] = \text{E}[\lambda_{ij}]$,   (5.9)

217   where E represents the expected value (i.e., mean) on the observed scale, β_0 is the mean value on

218   the latent scale (i.e. β_0 from the intercept-only model), σ_τ^2 is the total variance on the latent scale

11

219   (e.g., $\sigma_\alpha^2 + \sigma_e^2$ in Models 1 and 5, and σ_α^2 in Models 2-4 [2]; see also [27]). In fact, $\exp(\beta_0)$ gives

220   the median value of y_{ij} rather than the mean of y_{ij}, assuming a Poisson distribution. Thus, the use of

221    $\exp(\beta_0)$ will often overestimate σ_d^2, providing smaller estimates of R^2 and ICC, compared to when

222   using averaged y_{ij} (which is usually a better estimate of E[y_{ij}]). Quantitative differences between the

223   two approaches may often be negligible, but when λ is small, the difference can be substantial so

224   the choice of the method needs to be reported for reproducibility (Appendix S2). Our new

225   recommendation is to obtain λ via equation (5.8), which is the Poisson parameter averaged across

226   cluster-level parameters (λ_i for each individual in our example; [[17, 20, 28]]). Thus, obtaining λ via

227   equation (5.8) will be more accurate than estimating λ by calculating the average of observed

228   values although these two methods will give very similar or identical values when sampling is

229   balanced (i.e., observations are equally distributed across individuals and covariates). This

230   recommendation for obtaining λ also applies to negative binomial GLMMs (see Table 1).

231   6. Jensen's inequality and the 'second' delta method

232   A general form of equation (5.7) is known as Jensen's inequality, $g(\overline{x}) \leq \overline{g(x)}$ where g is a convex

233   function. Hence, the transformation of the mean value is equal to or larger than the mean of

234   transformed values (the opposite is true for a concave function; that is, $g(\overline{x}) \geq \overline{g(x)}$; [29]). In fact,

235   whenever the function is not strictly linear, simple application of the inverse link function (or back-

236   transformation) cannot be used to translate the mean on the latent scale into the mean value on the

237   observed scale. This inequality has important implications for the interpretation of results from

238   GLMMs, and also generalized linear models GLMs and linear models with transformed response

239   variables.

240   Although log-link GLMMs (e.g., Model 5) have an analytical solution, equation (5.8), this is not

241   usually the case. Therefore, converting the latent scale values into observation-scale values requires

242   simulation using the inverse link function. However, the delta method for bias correction can be

243   used as a general approximation to account for Jensen's inequality when using link functions or

244   transformations. This application of the delta method uses a second order Taylor series expansion

245   [18, 30]. A simple case of the delta method for bias correction can be written as:

246      $$\mathrm{E}[f(x)] \approx f(x) + 0.5\sigma_\tau^2 \frac{d^2}{dx^2} f(x) ,$$ (6.1)

247   where d^2/dx^2 is a second derivative with respect to the variable x and the other symbols are as in

248   equations (4.1) and (5.8). By employing this bias correction delta method (with

249      $d^2 \exp(x)/dx^2 = \exp(x)$), we can approximate equation (5.8) using the same symbols as in

250   equations (5.7)-(5.9):

251      $$\mathrm{E}[\lambda_{ij}] = \mathrm{E}[\exp(\beta_0)] \approx \exp(\beta_0) + 0.5\sigma_\tau^2 \exp(\beta_0)$$ (6.2)

252   The comparison between equation (5.8) (exact) and equation (6.2) (approximate) is shown in

253   Appendix S3. The approximation is most useful when the exact formula is not available as in the

254   case of a binomial GLMM with logit link (Model 6):

255      $$y_{ij} \sim \mathrm{binomial}(n_{ij},\ p_{ij},) ,$$ (6.3)

256      $$\mathrm{logit}(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + e_{ij} ,$$ (6.4)

257      $$\alpha_i \sim \mathrm{Gaussian}(0,\ \sigma_\alpha^2) ,$$ (6.5)

258      $$e_{ij} \sim \mathrm{Gaussian}(0,\ \sigma_e^2) ,$$ (6.6)

259   where y_{ij} is the number of 'success' in n_{ij} trials by the ith individual at the jth occasion (for binary

260   data, n_{ij} is always 1), p_{ij} is the underlying probability of success, and the other symbols are the same

261   as above. Binomial distributions have a mean of p and a variance of $np(1-p)$; (Table 2).

262   To obtain corresponding values between the latent scale and data (observation) scale, we need to

263   account for Jensen's inequality. The logit function used in binomial GLMMs combines of concave

13

264   and convex sections, which the delta method deals with efficiently. The overall intercept, β_0 on the

265   latent scale could therefore be transformed not with the inverse (anti) logit function

266   ($\mathrm{logit}^{-1}(x) = \exp(x)/(1+\exp(x))$), but with the bias-corrected delta method approximation. Given

267   that $d^2\mathrm{logit}^{-1}(x)/dx^2 = \exp(x)(1-\exp(x))/(1+\exp(x))^3$ in the case of the binomial GLMM with the

268   logit-link function, the approximation can be written as:

269     $$E[y_{ij}] = E[\mathrm{logit}^{-1}(\beta_0)] \approx \frac{\exp(\beta_0)}{1+\exp(\beta_0)} + 0.5\sigma_\tau^2 \frac{\exp(\beta_0)(1-\exp(\beta_0))}{(1+\exp(\beta_0))^3} \quad . \tag{6.7}$$

270   We can replace β_0 with any value obtained from the fixed part of the model (i.e. $\beta_0 + \sum \beta_h x_{hij}$).

271   McCulloch and colleagues [31] provide another approximation formula, which, by using our

272   notation, can be written as:

273     $$E[y_{ij}] \approx \mathrm{logit}^{-1}\left(\beta_0 - 0.5\sigma_\tau^2 \tanh\left(\frac{\beta_0(1+2\exp(-0.5\sigma_\tau^2))}{6} \right) \right) \tag{6.8}$$

274   Yet, another approximation proposed by Zeger and colleagues [32] can be written as:

275     $$E[y_{ij}] \approx \mathrm{logit}^{-1}\left(\beta_0 \left[\sqrt{1+\left(\frac{16\sqrt{3}}{15\pi}\right)^2 \sigma_\tau^2} \right]^{-1} \right). \tag{6.9}$$

276   This approximation, equation (6.9), uses the exact solution for the inverse probit function, which

277   can be written for a model like Model 6 but using the probit link: i.e.,

278     $\mathrm{probit}(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + e_{ij}$ in place of equation (6.4):

279     $$E[y_{ij}] = \mathrm{probit}^{-1}\left(\beta_0 \sqrt{1+\sigma_\tau^2}^{-1} \right). \tag{6.10}$$

280   A comparison between equations (6.7), (6.8) and (6.9) is also shown in Appendix S3 (it turns out

281   equation (6.8) gives the best approximation). Simulation will give the most accurate conversions

14

282   when no exact solutions are available. The use of the delta method for bias correction accounting

283   for Jensen's inequality is a very general and versatile approach that is applicable for any distribution

284   with any link function (see Appendix S3) and can save computation time. We note that the accuracy

285   of the delta method (both variance approximation and bias correction) depends on the form of the

286   function f, the conditions for and limitation of the delta method are described in the article by

287   Oehlert [30].

7.   Special considerations for binomial GLMMs

289   The observation-level variance σ_ε^2 can be thought of as being added to the latent scale on which

290   other variance components are also estimated in a GLMM (equations (3.2), (3.7), (3.12), (5.2) and

291   (6.4) for Models 2-6). Since the proposed R^2_{GLMM} and ICC $_{\text{GLMM}}$ are ratios between variance

292   components and their sums, we can show using the delta method that R^2_{GLMM} and ICC $_{\text{GLMM}}$

293   calculated via σ_ε^2 approximate to those of R^2 and ICC on the observation (original) scale (shown in

294   Appendix S4). In some cases, there exist specific formulae for ICC on the observation scale [2]. In

295   the past, we distinguished between ICC on the latent scale and on the observation scale [2]. Such a

296   distinction turns out to be strictly appropriate only for binomial distributions but not for Poisson

297   distributions (and probably also not for other non-Gaussian distributions). This is because the

298   property of what we have called the distribution-specific variance σ_d^2 for binomial distributions (e.g.

299     $\pi^2/3$ for binomial error distribution with the logit link function) is quite different from what we have

300   discussed as the observation-level variance σ_ε^2 although these two types of variance are related

301   conceptually (i.e., both represents variance due to non-Gaussian distributions with specific link

302   functions). Let us explain this further.

303   A binomial distribution with a mean of p (the proportion of successes) has a variance of $p(1–p)/n$

304   (the variance for the number of successes is $np(1–p)$; see Table 2). We find that the observation-

305   level variance is $1/(np(1–p))$ using the delta method on the logit-link function (see Table 2). This

15

306   observation-level variance $1/(np(1-p))$, or $1/(p(1-p))$ for binary data, is clearly different from the

307   distribution-specific variance $\pi^2/3$. As with the observation-level variance for the log-Poisson model

308   (which is $1/\lambda$ and changes with λ; note that we would have called $1/\lambda$ the distribution-specific

309   variance; [2, 3]), the observation-level variance of the binomial distribution changes as p changes

310   (see Appendix S5), suggesting these two observation-level variances ($1/\lambda$ and $1/(np(1-p))$) are

311   analogous while the distribution-specific variance $\pi^2/3$ is not. Further, the minimum value of

312   $1/(p(1-p))$ is 4, which is larger than $\pi^2/3 \approx 3.29$, meaning that the use of $1/p(1-p)$ in R^2 and ICC for

313   binary data will always produce larger values than those using $\pi^2/3$. Consequently, Browne and

314   colleagues [14] showed that ICC values (or variance partition coefficients, VPCs) estimated using

315   $\pi^2/3$ were higher than corresponding ICC values on the observation (original) scale using logistic-

316   binomial GLMMs (see also [33]). Note that they only considered binary data, i.e., $1/(np(1-p))$

317   where $n = 1$, because all proportion data can be rearranged as binary responses with a

318   grouping/clustering factor.

319   Then, what is $\pi^2/3$? Three common link functions in binomial GLMMs (logit, probit and

320   complementary log-log) all have corresponding distributions on the latent scale: the logistic

321   distribution, standard normal distribution and Gumbel distribution, respectively. Each of these

322   distributions has a theoretical variance, namely, $\pi^2/3$, 1 and $\pi^2/6$, respectively, which we previous

323   referred to as distribution-specific variances [2, 3] (Table 2). As far as we are aware, these

324   theoretical variances only exist for binomial distributions. The meaning of $1/(np(1-p))$, which is the

325   variance on the latent scale that approximates to the variance due to binomial distributions on the

326   observation scale is distinct from the meaning of $\pi^2/3$, which is the variance of the latent

327   distribution (i.e., the logistic distribution with the scale parameter being 1). The use of the

328   theoretical variance will almost always provide different values of R^2_{GLMM} and ICC $_{GLMM}$ from

329   those using the observation-level obtained via the delta method (see Appendix S5). This is because

330   the use of $\pi^2/3$ implicitly assumes all data sets have the same observation-level variance regardless

331   of mean (p) given the same number of trials (n). Therefore, we need distinguishing these theoretical

16

332   variances from the observation-level variance. R^2 and ICC values using the theoretical distribution-

333   specific variance might be rightly called the latent (link) scale (*sensu* [2]) whereas, as mentioned

334   above, R^2 and ICC values using the observation-level variance estimate the counterparts on the

335   observation (original) scale (cf. [25]).

8. Worked examples: revisting the beetles

337   In the following, we present a worked example by expanding the beetle dataset that was generated

338   for previous work [3]. In brief, the dataset represents a hypothetical species of beetle that has the

339   following life cycle: larvae hatch and grow in the soil until they pupate, and then adult beetles feed

340   and mate on plants. Larvae are sampled from 12 different populations ('Population'; see Figure 1).

341   Within each population, larvae are collected at two different microhabitats ('Habitat'): dry and wet

342   areas as determined by soil moisture. Larvae are exposed to two different dietary treatments

343   ('Treatment'): nutrient rich and control. The species is sexually dimorphic and can be easily sexed

344   at the pupa stage ('Sex'). Male beetles have two different color morphs: one dark and the other

345   reddish brown ('Morph', labeled as A and B in Figure 1). Sexed pupae are housed in standard

346   containers until they mature ('Container'). Each container holds eight same-sex animals from a

347   single population, but with a mix of individuals from the two habitats ($N_{[container]} = 120$; $N_{[animal]} =$

348   960).

349   We have data on five phenotypes, two of them sex-limited: (i) the number of eggs laid by each

350   female after random mating which we had generated previously using Poisson distributions (with

351   additive dispersion) and we revisit here for analysis with quasi-Poisson models (i.e. multiplicative

352   dispersion), (ii) the incidence of endo-parasitic infections that we generated as being negative

353   binomial distributed, (iii) body length of adult beetles which we had generated previously using

354   Gaussian distributions and that we revisit here for analysis with gamma distributions, (iv) time to

355   visit five predefined sectors of an arena (employed as a measure of exploratory tendencies) that we

356   generated as being gamma distributed, and (v) the two male morphs, which was again generated

17

357 with binomial distributions (for the details of parameter settings, see Table 3). We will use this

358 simulated dataset to estimate R^2_{GLMM} and ICC $_{\text{GLMM}}$.

359 All data generation and analyses were conducted in R 3.3.1 [10]. We used functions to fit GLMMs

360 from the three R packages: 1) the *glmmadmb* function from glmmADMB [34], 2) the *glmmPQL*

361 function from MASS [35], and 3) the *glmer* and *glmer.nb* functions from lme4 [36]. In Table 4, we

362 only report results from *glmmadmb* because this is the only function that can fit models with all

363 relevant distributional families. All scripts and results are provided as an electronic supplement

364 (Appendix S6). In addition, Appendix S6 includes an example of a model using the Tweedie

365 distribution, which was fitted by the *cpglmm* function from the cplm package [23]. Notably, our

366 approach for R^2_{GLMM} is kindly being implemented in the *rsquared* function in the R package

367 piecewiseSEM [37]. Another important note is that we often find less congruence in GLMM results

368 from the different packages than those of linear mixed-effects models, LMM. For example, GLMM

369 using the gamma error structure with the log-link function (Size and Exploration models),

370 *glmmadmb* and *glmmPQL* produced very similar results, while *glmer* gave larger R^2 and ICC values

371 than the former two functions (for more details, see Appendix S6; also see [38]). Thus, it is

372 recommended to run GLMMs in more than one package to check robustness of the results although

373 this may not always be possible.

374 In all the models, estimated regression coefficients and variance components are very much in

375 agreement with what is expected from our parameter settings (compare Table 3 with Table 4; see

376 also Appendix S6). When comparing the null and full models, which had 'sex' as a predictor, the

377 magnitudes of the variance component for the container effect always decrease in the full models.

378 This is because the variance due to sex is confounded with the container variance in the null model.

379 As expected, (unadjusted) ICC values from the null models are usually smaller than adjusted ICC

380 values from the full models because the observation-level variance (analogous to the residual

381 variance) was smaller in the full models, implying that the denominator of, for example, equation

382 (3.5) shrinks. However, the numerator also becomes smaller for ICC values for the container effect

18

383   from the parasite, size and exploration models so that adjusted ICC values are not necessarily larger

384   than unadjusted ICC values. Accordingly, adjusted $\text{ICC}_{[container]}$ is smaller in the parasite and size

385   models but not in the exploration model. The last thing to note is that for the morph models

386   (binomial mixed models), both R^2 and ICC values are larger when using the distribution-specific

387   variance rather than the observation-level variance, as discussed above (Table 4; see also Appendix

388   S4).

9. Alternatives and a cautonary note

390   Here we extend our simple methods for obtaining R^2_{GLMM} and ICC_{GLMM} for Poisson and binomial

391   GLMMs to other types of GLMMs such as negative binomial and gamma. We describe three

392   different ways of obtaining the observational-level variance and how to obtain the key rate

393   parameter λ for Poisson and negative binomial distributions. We discuss important considerations

394   which arise for estimating R^2_{GLMM} and ICC_{GLMM} with binomial GLMMs. As we have shown, the

395   merit of our approach is not only its ease of implementation but also that our approach encourages

396   researchers to pay more attention to variance components at different levels. Research papers in the

397   field of ecology and evolution often report only regression coefficients but not variance components

398   of GLMMs [3].

399   We would like to highlight two recent studies that provide alternatives to our approach. First, Jaeger

400   and colleagues [5] have proposed R^2 for fixed effects in GLMMs, which they referred to as $R^2_{\beta*}$ (an

401   extension of an R^2 for fixed effects in linear mixed models or R^2_{β} by Edwards and colleagues [39]).

402   They show that $R^2_{\beta*}$ is a general form of our marginal R^2_{GLMM}; in theory, $R^2_{\beta*}$ can be used for any

403   distribution (error structure) with any link function. Jaeger and colleagues highlight that in the

404   framework of $R^2_{\beta*}$, they can easily obtain semi-partial R^2, which quantifies the relative importance

405   of each predictor (fixed effect). As they demonstrate by simulation, their method potentially gives a

406   very reliable tool for model selection. One current issue for this approach is that implementation

407   does not seem as simple as our approach (see also [40]). We note that our R^2_{GLMM} framework could

408   also provide semi-partial R^2 via commonality analysis (see [41]), since unique variance for each

409   predictor in commonality analysis corresponds to semi-partial $R^{2;}$; [42].

410   Second, de Villemereuil and colleagues [25] have provided a framework with which one can

411   estimate exact heritability using GLMMs at different scales (e.g. data and latent scales). Their

412   method can be extended to obtain exact ICC values on the data (observation) scale, which is

413   analogous to, but not the same as, our ICC $_{\text{GLMM}}$ using the observation-level variance, σ_ε^2 described

414   above. Further, this method can, in theory, be extended to estimate R^2_{GLMM} on the data

415   (observation) scale. One potential difficulty is that the method of de Villemereuil and colleagues is

416   exact but that a numerical method is used to solve relevant equations so one will require a software

417   package (e.g., the QGglmm package; [25]). Relevantly, they have shown that heritability on the

418   latent scale does not need σ_d^2 (distribution-specific) but only need σ_e^2 (overdispersion variance),

419   which has interesting consequences in relation to our R^2_{GLMM} and ICC$_{\text{GLMM}}$ (we briefly describes

420   this possibility in Appendix S7; see also [40]).

421   Finally, we finish by repeating what we said at the end of our original R^2 paper [3]. Both R^2 and

422   ICC are indices that are likely to reflect only one or a few aspects of a model fit to the data and

423   should not be used for gauging the quality of a model. We encourage biologists use R^2 and ICC in

424   conjunctions with other indices like information criteria (e.g. AIC, BIC and DIC), and more

425   importantly, with model diagnostics such as checking for model assumptions, heteroscedasticity

426   and sensitivity to outliers.

Authors' contribuions

428   SN conceived ideas, and conducted analysis with discussions with HS. All developed the ideas

429   further, and contributed to writing and editing of the manuscript.

430 **Competing interests**

431   We have no competing interests.

432   **Funding**

436   **Acknowledgements**

440

441

References

[1] Lessells, C.M. & Boag, P.T. 1987 Unrepeatable repeatabilities - a common mistake. *Auk* **104**, 116-121.

[2] Nakagawa, S. & Schielzeth, H. 2010 Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* **85**, 935-956. (doi:10.1111/j.1469-185X.2010.00141.x).

[3] Nakagawa, S. & Schielzeth, H. 2013 A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* **4**, 133-142. (doi:10.1111/j.2041-210x.2012.00261.x).

[4] Johnson, P.C.D. 2014 Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods Ecol Evol* **5**, 944-946. (doi:10.1111/2041-210x.12225).

[5] Jaeger, B.C., Edwards, L.J., Das, K. & Sen, P.K. 2017 An R^2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics* **44**, 1086-1105. (doi:10.1080/02664763.2016.1193725).

[6] LaHuis, D.M., Hartman, M.J., Hakoyama, S. & Clark, P.C. 2014 Explained variance measures for multilevel models. *Organ Res Methods* **17**, 433-451. (doi:10.1177/1094428114541701).

[7] Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. 2009 Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* **24**, 127-135. (doi: 10.1016/J.Tree.2008.10.008).

[8] Bolker, B.M. 2008 *Ecological models and data in R*. Princeton, NJ, Princeton University Presss.

[9] Ver Hoef, J.M. & Boveng, P.L. 2007 Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* **88**, 2766-2772.

[10] R Development Core Team. 2016 R: A language and environment for statistical computing. (version 2.15.0 ed. Vienna, Austria, R Foundation for Statistical Computing.

465 [11] Snijders, T. & Bosker, R. 2011 *Multilevel Analysis: an Introduction to basic and advanced*

466 *multilevel modeling*. 2nd ed. London, Sage.

467 [12] Harrison, X.A. 2014 Using observation-level random effects to model overdispersion in count

468 data in ecology and evolution. *Peerj* **2**, e616. (doi:ARTN e61610.7717/peerj.616).

469 [13] Harrison, X.A. 2015 A comparison of observation-level random effect and Beta-Binomial

470 models for modelling overdispersion in Binomial data in ecology & evolution. *Peerj* **3**, e1114.

471 (doi:ARTN e111410.7717/peerj.1114).

472 [14] Browne, W.J., Subramanian, S.V., Jones, K. & Goldstein, H. 2005 Variance partitioning in

473 multilevel logistic models that exhibit overdispersion. *J R Stat Soc a Stat* **168**, 599-613.

474 [15] Efron, B. 1986 Double Exponential-Families and Their Use in Generalized Linear-Regression.

475 *J Am Stat Assoc* **81**, 709-721. (doi:10.2307/2289002).

476 [16] Gelfand, A.E. & Dalal, S.R. 1990 A note on overdispersed exponential-families. *Biometrika*

477 **77**, 55-64. (doi:10.2307/2336049).

478 [17] Gelman, A. & Hill, J. 2006 *Data analysis using regression and multilevel/hierarchical models*

479 Cambridge, Cambridge University Press.

480 [18] Ver Hoef, J.M. 2012 Who Invented the Delta Method? *Am Stat* **66**, 124-127.

481 (doi:10.1080/00031305.2012.687494).

482 [19] Powell, L.A. 2007 Approximating variance of demographic parameters using the delta method:

483 A reference for avian biologists. *Condor* **109**, 949-954. (doi: 10.1650/0010-

484 5422(2007)109[949:Avodpu]2.0.Co;2).

485 [20] Foulley, J.L., Gianola, D. & Im, S. 1987 Genetic Evaluation of Traits Distributed as Poisson-

486 Binomial with Reference to Reproductive Characters. *Theor. Appl. Genet.* **73**, 870-877. (doi:

487 10.1007/Bf00289392).

488 [21] Gray, B.R. & Burlew, M.M. 2007 Estimating trend precision and power to detect trends across

489 grouped count data. *Ecology* **88**, 2364-2372.

490   [22] Foster, S.D. & Bravington, M.V. 2013 A Poisson-Gamma model for analysis of ecological

491   non-negative continuous data. *Environ Ecol Stat* **20**, 533-552. (doi:10.1007/s10651-012-0233-0).

492   [23] Zhang, Y.W. 2013 Likelihood-based and Bayesian methods for Tweedie compound Poisson

493   linear mixed models. *Stat Comput* **23**, 743-757. (doi:10.1007/s11222-012-9343-7).

494   [24] Tempelman, R.J. & Gianola, D. 1999 Genetic analysis of fertility in dairy cattle using negative

495   binomial mixed models. *J Dairy Sci* **82**, 1834-1847.

496   [25] de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. 2016 General Methods for

497   Evolutionary Quantitative Genetic Inference from Generalized Mixed Models. *Genetics* **204**, 1281-

498   1294. (doi:10.1534/genetics.115.186536).

499   [26] Matos, C.A.P., Thomas, D.L., Gianola, D., Tempelman, R.J. & Young, L.D. 1997 Genetic

500   analysis of discrete reproductive traits in sheep using linear and nonlinear models .1. Estimation of

501   genetic parameters. *J Anim Sci* **75**, 76-87.

502   [27] Carrasco, J.L. 2010 A generalized concordance correlation coefficient based on the variance

503   components generalized linear mixed models for overdispersed count data. *Biometrics* **66**, 897-904.

504   (doi: 10.1111/J.1541-0420.2009.01335.X).

505   [28] Foulley, J.L. & Im, S. 1993 A marginal quasi-likelihood approach to the analysis of Poisson

506   variables with generalized linear mixed models. *Genet Sel Evol* **25**, 101-107. (doi:

507   10.1051/gse:19930107).

508   [29] Rao, C.R. 2002 *Linear statistical inference and its applications*. 2nd ed. ed. New York, John

509   Wiley & Sons.

510   [30] Oehlert, G.W. 1992 A note on the delta method. *Am Stat* **46**, 27-29. (doi: 10.2307/2684406).

511   [31] McCulloch, C.E., Searle, S.R. & Neuhaus, J.M. 2008 *Generalized, linear, and mixed models*.

512   2nd ed. ed. Hoboken, N.J., Wiley ; Chichester : John Wiley [distributor].

513   [32] Zeger, S.L., Liang, K.Y. & Albert, P.S. 1988 Models for Longitudinal Data - a Generalized

514   Estimating Equation Approach. *Biometrics* **44**, 1049-1060. (doi: 10.2307/2531734).

515   [33] Goldstein, H., Browne, W. & Rasbash, J. 2002 Partitioning variation in multilevel models.

516   *Understand Stat* **1**, 223-231.

517   [34] Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen,

518   A. & Sibert, J. 2012 AD Model Builder: using automatic differentiation for statistical inference of

519   highly parameterized complex nonlinear models. *Optim Method Softw* **27**, 233-249.

520   (doi:10.1080/10556788.2011.597854).

521   [35] Venables, W.N. & Ripley, B.D. 2002 *Modern applied statistics with S*. 4 ed. New York,

522   Springer.

523   [36] Bates, D., Machler, M., Bolker, B.M. & Walker, S.C. 2015 Fitting linear mixed-effects models

524   using lme4. *J Stat Softw* **67**, 1-48.

525   [37] Lefcheck, J.S. 2016 PIECEWISESEM: Piecewise structural equation modelling in R for

526   ecology, evolution, and systematics. *Methods Ecol Evol* **7**, 573-579. (doi:10.1111/2041-

527   210x.12512).

528   [38] Brooks, M.E., Kristensen, K.K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A.,

529   Skaug, H.J., Machler, M. & Bolker, B.M. 2017 Modeling zero-inflated count data with glmmTMB.

530   *bioRxiv* (doi: 10.1101/132753).

531   [39] Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F. & Schabenberger, O. 2008 An R^2

532   statistic for fixed effects in the linear mixed model. *Stat Med* **27**, 6137-6157. (doi:

533   10.1002/Sim.3429).

534   [40] Ives, A.R. 2017 R2s for correlated data: phylogenetic models, LMMs, and GLMMs. *bioRxiv*

535   (doi: 10.1101/144170).

536   [41] Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R. & Hamer, M. 2014

537   Using commonality analysis in multiple regressions: a tool to decompose regression effects in the

538   face of multicollinearity. *Methods Ecol Evol* **5**, 320-328. (doi: 10.1111/2041-210x.12166).

539   [42] Nimon, K.F. & Oswald, F.L. 2013 Understanding the results of multiple linear regression:

540   beyond standardized regression coefficients. *Organ Res Methods* **16**, 650-674.

541   (doi:10.1177/1094428113493929).

542

Table 1. The observation-level variance σ_ε^2 for the three distributional families: quasi-Poisson, negative binomial and gamma with the three different methods for deriving σ_ε^2: the delta method, log-normal approximation and the trigamma function, ψ_1.

Family	Distributional parameters	Mean (E[y]) Variance (var[y])	Link function	Delta method	Log-normal approximation	Trigamma function
Quasi-Poisson (QP)	$OP(\lambda,\ \omega)$	$E[y] = \lambda$	log	$\dfrac{\omega}{\lambda}$	$\ln\left(1 + \dfrac{\omega}{\lambda}\right)$	$\psi_1\left(\dfrac{\lambda}{\omega}\right)$
Poisson (when $\omega = 1$)	$\lambda > 0$ $\omega > 0$	$var[y] = \lambda\omega$	square-root	0.25ω	-	
Negative binomial (NB)	$NB(\lambda,\ \theta)$	$E[y] = \lambda$	log	$\dfrac{1}{\lambda} + \dfrac{1}{\theta}$	$\ln\left(1 + \dfrac{1}{\lambda} + \dfrac{1}{\theta}\right)$	$\psi_1\left(\left[\dfrac{1}{\lambda} + \dfrac{1}{\theta}\right]^{-1}\right)$
	$\lambda > 0$ $\theta > 0$	$var[y] = \lambda + \dfrac{\lambda^2}{\theta}$	square-root	$0.25\left(1 + \dfrac{\lambda}{\theta}\right)$	-	

27

Gamma	gmma(λ, ν)	$E[y] = \lambda$	log	$\dfrac{1}{\nu}$	$\ln\left(1+\dfrac{1}{\nu}\right)$	$\psi_1(\nu)$
	$\lambda > 0$	$\mathrm{var}[y] = \dfrac{\lambda^2}{\nu}$	inverse	$\dfrac{1}{\nu\lambda^2}$	-	
	$\nu > 0$		(reciprocal)			
Gamma (alternative parameterization)	gamma(ν, κ)	$E[y] = \dfrac{\nu}{\kappa}$	log	$\dfrac{1}{\nu}$	$\ln\left(1+\dfrac{1}{\nu}\right)$	$\psi_1(\nu)$
	$\nu > 0$	$\mathrm{var}[y] = \dfrac{\nu}{\kappa^2}$	inverse	$\dfrac{\kappa^2}{\nu^3}$	-	
	$\kappa > 0$		(reciprocal)			

$\mathrm{var}[\ln(x)] = \psi_1(\nu) = \sum_{n=1}^{\infty} 1/(\nu+n)$ when x follows gamma distribution. In the R environment, the function, *trigamma* can be used to obtain $\psi_1(\nu)$; also note that

ν is known as a shape parameter while κ is as a rate parameter in gamma distribution.

Table 2. The distribution-specific (theoretical) variance σ_d^2 and observation-level variance σ_ε^2 using the delta method for binomial (and

Bernoulli) distributions; note that only one of them should be used for obtaining R^2 and ICC.

Family	Distributional parameters, mean & variance	Link name	Link function	Theoretical (distribution-specific) variance	Observation-level variance (min. values and corresponding p given $n = 1$)
Binomial (Bernoulli; $n = 1$)	binomial(n, p) $0 < p < 1$ $n >= 1$ (integers) E[y] = np var[y] = $np(1-p)$ var[y/n] = $p(1-p)$/n	logit	$\ln\left(\dfrac{p}{1-p}\right)$	$\dfrac{\pi^2}{3} \approx 3.29$ (logistic distribution)	$\dfrac{1}{np(1-p)}$ (min = 4; $p = 0.5$)
		probit ($\Phi(p)$)	$\sqrt{2}\,\text{erf}^{-1}(2p-1)$	1 (standard normal distribution)	$2\pi n^{-1} p(1-p)\left(\exp\left[\text{erf}^{-1}(2p-1)\right]^2\right)^2$ (min ~ 1.57; $p = 0.5$)

29

cloglog			
(complimentary	$\ln(-\ln(1-p))$	$\dfrac{\pi^2}{6} \approx 1.65$	$\dfrac{p}{n\left(\ln(1-p)\right)^2 (1-p)}$
log-log)		(Gumbel distribution)	(min ~ 1.54; p ~ 0.8; ~ 2.08; $p = 0.5$)

'erf^{-1}' is the inverse of the Gauss error function, which is often denoted as 'erf'.

Table 3. Parameter settings of regression coefficients (*b*) and variance components (σ^2) for five data sets: 1) fecundity, 2) endoparasite, 3) size, 4) exploration and 5) morph; all parameters are set on the latent scale apart from the size data (see below).

Response	Intercept (*b*)	Sex (*b*)	Treatment (*b*)	Habitat (*b*)	Population (σ^2)	Container (σ^2)	Overdispersion (σ^2)
Fecundity: the number of eggs per female	1.1	-	0.5	0.1	0.4	0.05	0.1
Parasite: the number of endoparasites per individual	1.8	-2	-0.8	0.7	0.5	0.8	-
Size: the body length of an individual*	15	-3	0.4	0.15	1.3	0.3	1.2
Exploration: the time taken visiting five sectors for an individual	4	-1	2	-0.5	0.2	0.2	-
Morph colour morph of a male	-0.8	-	0.8	0.5	1.2	0.2	-

*Data for the six sets of models were simulated on the normal (Gaussian) scale but analysed assuming a gamma error structure with the log link so that estimations of these parameters will be on the log scale; note the overdispersion variance for this data is the residual variance.

Table 4. Mixed-effects model analysis of a simulated dataset estimating variance components and regression slopes for nutrient manipulations on fecundity, endoparasite loads, body length, exploration levels and male morph types; $N_{[population]}$=12, $N_{[container]}$=120 and $N_{[animal]}$=960 ($N_{[male]}$= $N_{[female]}$ = 480).

Model name	Fecundity models (log-link)		Parasite models (log-link)		Size models (log-link)		Exploration models (log-link)		Morph models (logit-link)	
	Quasi-Poisson mixed models		Negative binomial mixed models		Gamma mixed models		Gamma mixed models		Binomial (binary) mixed models	
	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model
Fixed effects	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*
	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]
Intercept	1.630	1.261	0.766	1.752	2.682	2.737	4.752	4.056	-0.108	-0.740
	[1.379, 1.882]	[0.989, 1.532]	[0.330, 1.202]	[1.282, 2.223]	[2.616, 2.689]	[2.699, 2.775]	[4.555, 4.949]	[3.842, 4.269]	[-0.718, 0.501]	[-1.450, -0.030]
Treatment (experiment)	-	0.491	-	-0.768	-	0.033	-	2.007	-	0.840
		[0.391, 0.591]		[-0.870, -0.667]		[0.023, 0.044]		[1.965, 2.050]		[0.422, 1.258]

33

Habitat (wet)	-	0.152	-	0.700	-	0.009	-	-0.560	-	0.414
		[0.055, 0.249]		[0.599, 0.801]		[-0.001, 0.019]		[-0.603, -0.518]		[0.002, 0.826]
Sex (male)	-	-	-	-2.198	-	-0.213	-	-1.105	-	-
				[-2.511, -1.884]		[-0.230, -0.196]		[-1.256, -0.955]		-
Random effects	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2
Population	0.178	0.187	0.375	0.541	0.0026	0.0039	0.071	0.104	1.002	1.111
Container	0.042	0.059	1.976	0.613	0.0140	0.0014	0.364	0.163	0.136	0.186
Observation-level (Distribution-specific)	0.477	0.349	0.873	0.397	0.0069	0.0064	1.664	0.118	4.010 (3.290)	4.010 (3.290)
Fixed factors	-	0.066	-	1.479	-	0.0116	-	1.393	-	0.220
$R^2_{\text{GLMM}(m)}$	-	9.96%	-	48.50%	-	49.54%	-	78.34%	-	3.98% (4.57%)

$R^2_{\text{GLMM}(c)}$	-	46.95%	-	86.33%	-	72.52%	-	93.34%	-	27.46% (31.55%)
ICC[Population]	25.33%	31.30%	11.53%	34.44%	11.38%	33.17%	3.40%	26.94%	19.48% (22.64%;)	20.95% (24.23%)
ICC[Container]	5.94%	9.79%	60.80%	39.02%	59.57%	12.37%	17.34%	42.34%	2.64% (3.07%;)	3.50% (4.05%)
AIC	2498.8	2412.3	4342.6	3920.5	3379.9	3139.5	11223.8	9004.3	605.5	589.6

95 % CI (confidence intervals) were calculated by the *confint* function in lme4. The observation-level variance was obtained by using the trigamma function. In the Morph models, both the observation-level variance and (theoretical) distribution-specific variance were used; note that ones in brackets use the distribution-specific variance for R^2 and ICC. ICC[Container] is not a typical 'repeatability' but the proportion of variance due to the container effect beyond the population variance.

Figure legends

Figure 1. A schematic of how hypothetical datasets are obtained (see the main text for details).