

iccCounts: An R Package to Estimate the Intraclass Correlation Coefficient for Assessing Agreement with Count Data

by Josep L. Carrasco

Abstract The intraclass correlation coefficient (ICC) is a widely used index to assess agreement with continuous data. The common approach for estimating the ICC involves estimating the variance components of a linear mixed model under assumptions such as linearity and normality of effects. However, if the outcomes are counts these assumptions are not met and the ICC estimates are biased and inefficient. In this situation, it is necessary to use alternative approaches that are better suited for count data. Here, the *iccCounts* R package is introduced for estimating the ICC under the Poisson, Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial distributions. The utility of the *iccCounts* package is illustrated by three examples that involve the assessment of repeatability and concordance with count data.

1 Introduction

Repeated measurements are often collected and hierarchically structured in clusters (commonly, in subjects). These repeated measurements can be interchangeable within subjects (i.e. they are replicates). This case is often known as the evaluation of repeatability (Nakagawa and Schielzeth 2010) or intra-rater reliability (DeVet et al. 2011). Moreover, the repeated measurements may be structured (not interchangeable) because they were obtained under different experimental conditions, involving different methods or observers. In this case the analysis of agreement is often known as concordance analysis, method comparison analysis (Choudhary and Nagaraja 2017) or inter-rater reliability (DeVet et al. 2011).

Whatever the structure of the repeated measurements, the intraclass correlation coefficient (ICC) is a common index used to assess agreement with continuous data (Fleiss 1986; Carrasco and Jover 2003). The general definition of the ICC is the ratio of the between-clusters variance to the total variance. However, the appropriate ICC has to be defined to afford the different variance components that are involved in the total variance besides the between-clusters variance. These variance components are typically estimated by means of a linear mixed model with common assumptions: that there is linearity between the outcome expectation and the effects (cluster, method,...); and that the random effects and the random error follow normal distributions. Currently, there are several R packages that estimate the ICC under the Normality assumption for assessing repeatability or concordance (Wolak, Fairbairn, and Paulsen 2012; Carrasco et al. 2013; Stoffel, Nakagawa, and Schielzeth 2017).

However, if the outcomes are counts, such assumptions are not met and the ICC estimates are biased and inefficient (Carrasco and Jover 2005). In this situation, it is necessary to use alternative approaches that are better suited to the properties of count data. The methodology for estimating the ICC for non-normal distributions using generalized linear mixed models (GLMM), and in particular for count data, was developed in Carrasco (2010). The ICC is therefore estimated by the variance components from the appropriate GLMM. The cluster random effect is still distributed as a Normal distribution but the within-cluster variability is assumed to follow a probability distribution function for counts. Stoffel, Nakagawa, and Schielzeth (2017) introduced the *rptR* package which can be used to estimate the ICC assuming a Poisson model for the within-subjects variability.

In the *iccCounts* package introduced here, besides the Poisson distribution, other models as the Negative Binomial, the Zero-Inflated Poisson and the Zero-Inflated Negative Binomial are also considered. These models are useful when overdispersion arises in the Poisson model. Overdispersion means that the variability assumed by the model is lower than that from the data. Therefore, the within-subjects variability and, by extension, the total variance are underestimated and the ICC and its standard error are biased. Thus, the validity of the ICC estimate is closely linked to the validity of the model, so that a goodness-of-fit (GOF) analysis of the model must be performed.

The article is structured as follows: the *Methodology background* section introduces the definition of the ICC, their expressions depending on the model GLLM chosen, some inferential aspects and the validation approach of the GLMM; the issues of the package are described in *Package description* section; in *Examples* section three examples are introduced. Two of them are cases of the repeatability setting whereas the remaining one shows the case of a concordance setting. Finally, the main contributions are summarized in the *Conclusion* section.

2 Methodology background

Experimental design

As mentioned in the introduction, the experimental design depends on the aim of the study: concordance or repeatability. In the case of a concordance study, a sample of n subjects are measured m times by k methods. In this setting, the aim is to analyse the degree of concordance of the measurement methods when assessing the subjects. Note that within-subjects repeated measurements are not interchangeable because they belong to one specific method. It is worthy to noting that the term “methods” is the conventional way to describe the experimental condition of the repeated measurements across subjects in this context, but it might be referred to differently name depending on the context. In this setting, Y_{ijk} stands for the k -th reading made by the j -th method on the i -th subject, with $i = 1, \dots, n$, $j = 1, \dots, m$ and $k = 1, \dots, s$. Hence, there will be three variance components to consider when assessing agreement among the repeated measurements: between-subjects, between-methods and random error variabilities.

In a repeatability study, a sample of n subjects are measured m times. In this case the repeated measurements share the same experimental condition across subjects, therefore they can be considered as interchangeable. Thus, in this setting Y_{ik} stands for the k -th reading made on the i -th subject, with $i = 1, \dots, n$, and $k = 1, \dots, s$. In this case, only two variance components are involved in the evaluation of the agreement: between-subjects and random error variabilities.

Generalized linear mixed model

The estimation of the variance components is carried out by means of generalized linear mixed models (GLMM). The GLMM for the concordance setting (considering subjects and methods effects) is defined as follows:

- Let α_i and β_j be the subjects and methods random effects respectively, with $i = 1, \dots, n$, $j = 1, \dots, m$, that follow Normal distributions with mean 0 and variance σ_α^2 and σ_β^2 . Although the method effect could be a fixed effect by design, when defining the agreement index it is convenient to consider it as a random effect to account for the systematic differences between observers as a source of disagreement (Fleiss 1986; Carrasco and Jover 2003).
- The conditional distribution of Y_{ijk} given α_i and β_j , $f(Y_{ijk}|\alpha_i, \beta_j)$, is a probability density function from the exponential family.
- The conditional mean of Y_{ijk} given α_i and β_j is

$$\mu_{ij} = E(Y_{ijk}|\alpha_i, \beta_j) = g^{-1}(\lambda_i + \alpha_i + \beta_j). \quad (1)$$

where g is called the link function. Here, λ_i is the linear combination of the mean modifying covariates for the i -th subject. Furthermore, the conditional variance of Y_{ijk} given α_i and β_j is defined as $Var(Y_{ijk}|\alpha_i, \beta_j) = \phi h(\mu_{ij})$ where ϕ is the dispersion parameter and $h(\cdot)$ is variance function.

If the methods effect is removed, the GLMM for the repeatability setting is obtained. Thus, depending on the nature of the data the appropriate conditional probability model and link function must be chosen. When analysing count data, the logarithm is commonly used as a link function and models such as Poisson or Negative Binomial are considered.

Intraclass correlation coefficient for count data

The intraclass correlation coefficient (ICC) is calculated as:

$$ICC = \frac{Cov(Y_{ijk}, Y_{ij'k'})}{Var(Y_{ijk})}. \quad (2)$$

where the $Cov(Y_{ijk}, Y_{ij'k'})$ is the marginal (over subjects and observers) covariance between any pair of data from the same subject, whereas $Var(Y_{ijk})$ is the marginal variance of data.

Furthermore, the marginal variance and covariance can be developed as functions of the GLMM parameters (Carrasco 2010):

$$ICC = \frac{Cov(\mu_{ij}, \mu_{ij'})}{Var(\mu_{ij}) + E(\phi h(\mu_{ij}))} \quad (3)$$

This result allows the ICC to be generalized to any distribution fitted with a GLMM.

Carrasco (2010) developed the ICC for Poisson and Negative Binomial distributions, the latter with variance increasing quadratically with the mean (NegBin2), $Var(Y_{ijk}|\alpha_i, \beta_j) = \mu_{ij}(1 + r\mu_{ij})$ (Table 1). A Negative Binomial model with variance increasing linearly with the mean is also considered (NegBin1) (Brooks et al. 2017; Hardin and Hilbe 2007), $Var(Y_{ijk}|\alpha_i, \beta_j) = \mu_{ij}(1 + r)$. It is worth noting that NegBin1 does not belong to the exponential family (Hardin and Hilbe 2007), therefore this model would not be a proper GLMM. However, it can still be useful to model count data that show overdispersion in a Poisson model.

Additionally, it is possible to define the ICC for the cases of zero inflated models (Table 1). Let's define B_{ijk} as a Bernoulli variable that takes a value of 1 if the reading k on subject i and method j is a structural zero with probability π and 0 otherwise. The observed data, X_{ijk} is the result of $X_{ijk} = Y_{ijk}(1 - B_{ijk})$, where Y_{ijk} is the count variable as defined before. The marginal covariance and variance of X_{ijk} are:

$$Cov(X_{ijk}, X_{ij'k'}) = Cov(Y_{ijk}, Y_{ij'k'}) (1 - \pi)^2 \quad (4)$$

$$Var(X_{ijk}) = Var(Y_{ijk}) (1 - \pi) + E^2(Y_{ijk}) \pi (1 - \pi) \quad (5)$$

and the general expression of the ICC for zero-inflated data becomes:

$$ICC = \frac{Cov(Y_{ijk}, Y_{ij'k'}) (1 - \pi)}{Var(Y_{ijk}) (1 - \pi) + E^2(Y_{ijk}) \pi} \quad (6)$$

where π stands for the probability of excess of zeros.

Notice that ICCs appearing in Table 1 are for the concordance setting where σ_β^2 stands for the variability between methods. The ICCs for the repeatability setting are obtained just removing σ_β^2 from the equations, i.e. by setting $\sigma_\beta^2 = 0$.

Model	ICC	θ	Model	ICC	θ
Poisson	$\frac{\mu(e^{\sigma_\alpha^2}-1)}{\mu(e^{\sigma_\alpha^2+\sigma_\beta^2}-1)+1}$	$(\mu, \sigma_\alpha^2, \sigma_\beta^2)$	ZIP	$\frac{\mu(e^{\sigma_\alpha^2}-1)(1-\pi)}{\mu(e^{\sigma_\alpha^2+\sigma_\beta^2}-1)+1+\mu\pi}$	$(\mu, \sigma_\alpha^2, \sigma_\beta^2)$
NegBin1	$\frac{\mu(e^{\sigma_\alpha^2}-1)}{\mu(e^{\sigma_\alpha^2+\sigma_\beta^2}-1)+r+1}$	$(\mu, \sigma_\alpha^2, \sigma_\beta^2, r)$	ZI-NegBin1	$\frac{\mu(e^{\sigma_\alpha^2}-1)(1-\pi)}{\mu(e^{\sigma_\alpha^2+\sigma_\beta^2}-1)+1+r+\mu\pi}$	$(\mu, \sigma_\alpha^2, \sigma_\beta^2, r, \pi)$
NegBin2	$\frac{\mu(e^{\sigma_\alpha^2}-1)}{\mu((r+1)e^{\sigma_\alpha^2+\sigma_\beta^2}-1)+1}$	$(\mu, \sigma_\alpha^2, \sigma_\beta^2, r)$	ZI-NegBin2	$\frac{\mu(e^{\sigma_\alpha^2}-1)(1-\pi)}{\mu((r+1)e^{\sigma_\alpha^2+\sigma_\beta^2}-1)+1+\mu\pi}$	$(\mu, \sigma_\alpha^2, \sigma_\beta^2, r, \pi)$

Table 1: ICC expressions. NegBin1 and NegBin2 are the Negative Binomial models with variance increasing linearly and quadratically with the mean respectively; ZI are the zero-inflated models. θ stand for parameters involved in ICC: μ stands for the marginal mean; σ_α^2 is the between-subjects variance; σ_β^2 is the between-methods variability; r is the Negative Binomial's dispersion parameter; and π is the probability of excess of zeros.

Estimation of ICC

The estimation of the ICC involves estimating the GLMM parameters. However, maximum likelihood approach is not straightforward because there is no closed analytical expression for the marginal likelihood besides the Normal case (linear mixed model). Thus, it is necessary to apply numerical methods to approximate the marginal likelihood and to obtain maximum likelihood estimates (Bolker et al. 2009). With regards to the standard error of the ICC, let θ be the GLMM parameters involved in the ICC expression (see Table 1) and Σ the variance-covariance matrix of θ . The asymptotic standard error can be estimated by applying the delta method (Hoef 2012):

$$\text{Var}(\text{ICC}) \approx \Delta' \Sigma \Delta \quad (7)$$

where Δ stand for the vector containing the derivatives of ICC respect to θ . Confidence intervals for the ICC are based on asymptotic Normal distribution using the inverse hyperbolic tangent function or Fisher's Z-transformation (Carrasco and Jover 2003).

Validation

The goodness-of-fit (GOF) analysis can be carried out by the computation of randomized quantile residuals (RQR) (Dunn and Smyth 1996; C. Feng, Li, and Sadeghpour 2020). Briefly, the GOF analysis involve the comparison of the RQR from the original data to those obtained by simulation under the fitted model. Simulations of counts based on the fitted model are generated and the model is refitted to each simulated dataset. Using the simulated RQR, envelopes are built as the appropriate quantiles (in relation to the level of significance) of RQR from the refitted models. If the model fits correctly the data it is expected that the original RQR will completely lie within the simulated envelopes.

Additionally, dispersion as well as zero-inflation can be checked by comparing the dispersion and proportion of zeros from the simulated data to those from the original data. Thus, tests for dispersion and zero inflation are carried out by comparing the RQR dispersion and the number of zeros from the original model and data to those from the refitted models and simulated data.

3 Package description

The main function in the **iccCounts** package is `icc_counts` which estimates the ICC under different models for count data. The argument `data` identifies the data set to be analysed. This data set has to be a `data.frame` object with at least two columns: outcome and subject identifier (arguments `y` and `id` respectively).

In the case of estimating the ICC for the concordance setting, a third column with the method identifier must be provided (the argument `met`). The argument `type` is used to identify the setting in which the ICC should be estimated. Valid values are: `rep` (default) for the repeatability setting; and `con` for the concordance setting. The repeatability setting requires that repeated measurements are interchangeable within subjects. This means the experimental conditions of the measurements are the same (replicates), and they come from the same probability distribution function (conditioned to subjects). On the other hand, in the concordance setting the repeated measurements are not interchangeable because their experimental conditions are different, and therefore their probability distribution function, conditioned to subjects, is different (commonly in the mean).

The argument `fam` is used to identify the within-subjects probability model. Valid options are: `poisson` (default) for Poisson model; `nbinom1` and `nbinom2` for Negative Binomial model with variance increasing linearly and quadratically with the mean respectively; `zip` for zero-inflated Poisson model; `zinb1` and `zinb2` for zero-inflated Negative Binomial model with variance increasing linearly and quadratically with the mean.

Once the appropriate setting and model have been chosen, the GLMM is estimated by maximum likelihood via Laplace approximation using the **glmmTMB** package (Brooks et al. 2017). The output of the `icc_counts` function is an object of class `iccc` which is a list with the following components: `model` which contains the generalized linear mixed model estimates; `ICC` which includes the ICC estimate and its standard error and confidence interval; and `varcomp` with the variance components and parameters related to the ICC. Finally, the function `GOF_check` runs the goodness of fit (GOF) analysis of the GLMM fitted to data. This function has three arguments: `x` to denote the `iccc` object to apply the GOF analysis; the `nsim` argument that stands for the number of simulations to run which default value is set to 100; and the `alpha` argument to set the level of significance.

The output of `GOF_check` is an object of class `GOF` which is a list with the following components: `plot_env`, a plot of RQR envelopes with the original RQR; `plot_var`, a plot of the simulated RQR dispersion; `plot_zi`, a plot of the count of zeros in the simulated datasets; `res_var`, the dispersion of RQR from the original sample; `pval_var`, the proportion of simulated RQR dispersion that are greater than the original dispersion that can be interpreted as a simulated P-value to check the goodness of fit on dispersion; `zero_count`, the count of zeros in the original sample; and `pval_zi`, the proportion of simulated zero count that are greater than that of the original sample. It can be interpreted as a simulated P-value to check the hypothesis of zero-inflation. The plots in the list are objects of class `ggplot`, hence users may change the plot themes or add modifications to the components.

Additionally, to describe the differences among the repeated measurements from the same subjects, the function `plot_BA` draws the Bland-Altman plot (Bland and Altman 1995). The difference between

each pair of data from the same subject is represented on the y-axis. The mean of data from the same subject is represented on the x-axis. Additionally, a bar plot with the proportions of differences can be drawn. This plot is a useful way to describe the differences when the range of observed values is small relative to the number of observations (Smith, Ma, and Stafford 2010).

The arguments of *plot_BA* function are: *data*, a data frame containing at least the columns of the outcome and subject's identifier; *y*, a character string indicating the name of the outcome column in the data set; *id* a character string indicating the name of the subjects column in the data set; *rm*, a character string indicating the name of the column that stands for the repeated measurements in the data set. This argument is only needed to identify the differences; *type*, argument used to choose the plot to be drawn. Valid values are: *BA* (default) for the Bland-Altman plot; and *bars* for the bar plot of the differences. Besides the plots, the function provides a dataframe object that contains the data used to generate the plot.

4 Examples

The package includes three real data sets as examples that covers the repeatability and concordance settings.

Sparrow fledglings paternity example

In the Sparrow fledglings paternity example, the incidence of extra-pair paternity (EPP) was monitored over 3 breeding seasons in a sparrow colony in Lundy, an island off the southwest coast of England (Schroeder et al. 2012). One of the aims of the study was to assess the repeatability of counts of fledglings that a male had in every breeding season. Thus, the repeated measurements are assumed to be exchangeable replicates. However, the means of the Social variable by year seem to differ:

```
library(iccCounts)
library(dplyr)
EPP %>% group_by(Year) %>% summarize(Mean=mean(Social),SD=sd(Social))

#> # A tibble: 3 x 3
#>   Year Mean SD
#>   <int> <dbl> <dbl>
#> 1  2003  3.19  3.10
#> 2  2004  2.53  2.21
#> 3  2005  4.5   2.79
```

In case these means were significantly different the repeated measurements could not be considered as exchangeable, and consequently the differences among the means of the repeated measurements should be included in the agreement index. This led us to the concordance setting considering Year as the methods effect.

The first model to consider is the Poisson model. The default option in the *icc_counts* function is the Poisson model, and so it is necessary to specify the name of the data set, the count variable (Social), the subjects identifier (id), the methods variable (Year), and the concordance setting (type="con").

```
EPP_P<-icc_counts(EPP,y="Social",id="id",met="Year",type="con")
ICC(EPP_P)

#>           ICC      SE ICC 95% CI LL 95% CI UL
#> [1,] 0.5284404 0.0866857 0.3383706 0.6770822

VarComp(EPP_P)

#>      mu      BSVar      BMVar
#> 3.172783 0.4002965 0.0798841
```

The function *ICC* applied to the *iccc* object shows that the ICC estimate is 0.53 (95% confidence interval: 0.34 - 0.68). Moreover, the function *VarComp* gives the parameters involved in the ICC estimator, which in this case are the overall mean, the between-subjects variance and the between-methods variability (the two latter in log-scale).

However, as mentioned in the previous section, the validity of the ICC estimate is linked to the validity of the model. The function *GOF_check* is applied to the *iccc* object to run the simulations and to compute the RQR. The random seed is set for the sake of the reproducibility of the example.

```
set.seed(100)
EPP_P.gof <- GOF_check(EPP_P)

plot(EPP_P.gof)
```

Figure 1a shows the plot of RQR with envelopes generated by simulation. Points on the plot stand for the RQR from the original sample. Notice that a substantial number of points lie outside the envelopes, indicating the fit of the model is unsuitable. The next plot (Figure 1c) shows the density of the RQR variances computed in the simulated samples. The RQR variance from the initial sample is 2.15 (shown inside the square) which is extreme compared to those from the simulations. Indeed, the proportion of simulated variances that are higher than that from the initial sample can be interpreted as a p-value generated by Monte Carlo simulation. This p-value is shown by applying the function *DispersionTest* to the *GOF* object.

```
DispersionTest(EPP_P.gof)

#>      S      P_value
#> 2.077549 0.00990099
```

Additionally, the Social variable has a considerable proportion of zero values (26.4%), and so the excess of zeros could be the cause of the unsuitable fitting of the Poisson model. To check this hypothesis, the third plot generated shows the proportion of zeros in the simulated data sets (Figure 1e). The count of zeros in the sample is 51, which exceeds the expected count under the Poisson model. Again, the proportion of simulated zero counts that are higher than that from the initial sample can be interpreted as a p-value generated by Monte Carlo simulation. This p-value is obtained by applying the function *ZeroTest* to the *GOF* object.

```
ZeroTest(EPP_P.gof)

#> Count      P_value
#>    51 0.00990099
```

Thus, it is necessary to use a model able to provide a proportion of zeros higher than that expected under the Poisson assumption. This model could be the Zero-Inflated Poisson (ZIP) model.

```
EPP_ZIP<-icc_counts(EPP,y="Social",id="id",met="Year",type="con",fam="zip")
ICC(EPP_ZIP)

#>      ICC      SE ICC 95% CI LL 95% CI UL
#> [1,] 0.0477628 0.0362002 -0.02331 0.1183553

VarComp(EPP_ZIP)

#>      mu      BSVAR      BMVar      Pi
#> 4.487117 0.033328 0.0328427 0.2446678
```

In this case, the ICC is much lower than in the Poisson model (0.05, 95% CI: -0.02, 0.12) indicating a non-significant ICC. The ICC components are the same as those in the Poisson case (with different values) plus the proportion of excess of zeros (0.24). Next step is to check whether the model correctly fits the data.

```
set.seed(100)
EPP_ZIP.gof <- GOF_check(EPP_ZIP)

plot(EPP_ZIP.gof)
```

Figure 1b shows the model to be appropriate because all the RQR are within the envelopes. Additionally, the dispersion and the proportion of zeros from the initial sample are within the values expected under the ZIP model (Figures 1d and 1f). This fact can be also verified by verifying that dispersion and excess of zeros tests are non significant.

```
DispersionTest(EPP_ZIP.gof)
```

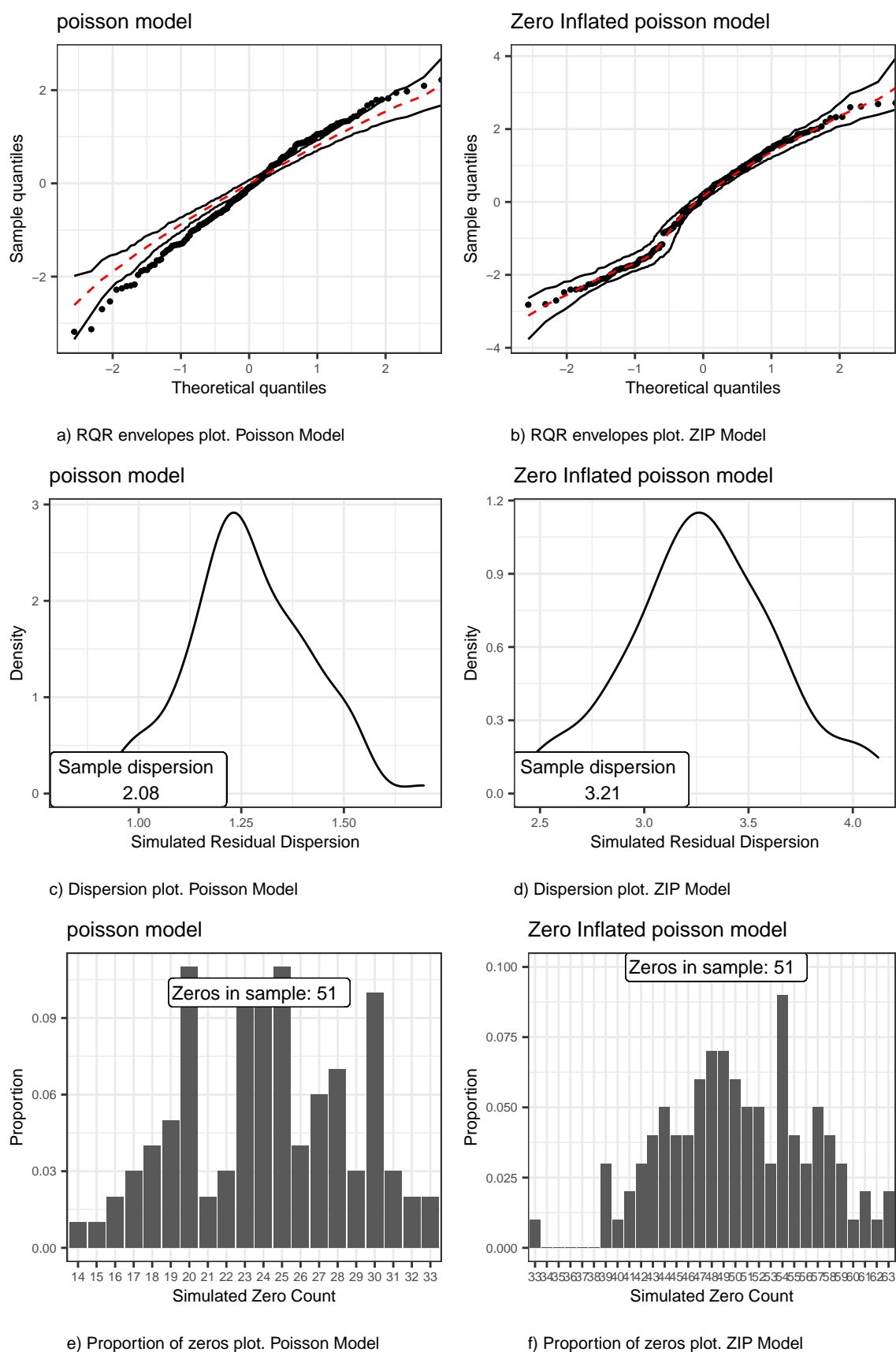


Figure 1: Goodness of fit for Sparrow fledglings paternity example. The Randomized Quantile Residuals (RQR) and counts of zeros of original data are compared to those from simulated data under the fitted model. The plots shown are RQR with envelopes, dispersion of RQR and count of zeros. Left column shows results for Poisson model while the plots for Zero Inflated Poisson (ZIP) model are on right column.

```
#>      S    P_value
#> 3.214152 0.5643564
```

```
ZeroTest(EPP_ZIP.gof)
```

```
#> Count    P_value
#>    51 0.4752475
```

Thus, the ZIP model fits the data appropriately. The next step is to check if the differences in means between years are significant and the concordance setting is therefore justified. With this aim let us apply the function `icc_counts` to the ZIP model but in the repeatability setting.

```
EPP_ZIP_0<-icc_counts(EPP,y="Social",id="id",fam="zip")
```

The component `model` in the `iccc` object is an object of class `glmmTMB` that contains the generalized linear mixed model fit. The `anova` method applied to the model objects gives a comparison of deviances and a likelihood ratio test:

```
anova(EPP_ZIP$model,EPP_ZIP_0$model)
```

```
#> Data: data
#> Models:
#> EPP_ZIP_0$model: y ~ (1 | id), zi=~1, disp=~1
#> EPP_ZIP$model: y ~ met + (1 | id), zi=~1, disp=~1
#>      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
#> EPP_ZIP_0$model  3 847.63 857.42 -420.82  841.63
#> EPP_ZIP$model    5 836.35 852.67 -413.18  826.35 15.279      2 0.0004811 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result of the comparison confirms a significant difference between the yearly means, and the convenience of the concordance setting.

Next, let us apply the function `plot_BA` to generate the Bland-Altman plot and the bar plot of the differences in EPP between years. The plots are shown in Figures 4a and 4b.

```
EPP.BA<-plot_BA(EPP,y="Social",id="id",rm="Year") # Bland-Altman plot
```

```
plot_BA(EPP,y="Social",id="id",type="bars") # Bar plot
```

It can be seen that the magnitude of the differences grows as the mean increases. This heteroscedastic pattern is expected in counts because of the relation between the within-subjects variance and mean. Furthermore, we can compute some descriptive statistics of the differences:

```
summary(EPP.BA$data$Diff)
```

```
#>   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#> -7.0000 -1.0000  1.0000  0.9737  3.0000 10.0000
```

```
quantile(EPP.BA$data$Diff,probs=c(0.05,0.95))
```

```
#>  5% 95%
#> -4   6
```

Briefly, the mean of the differences between years is 0.97 fledglings, and the median is 1 fledgling. Ninety percent of the differences are between -4 and 6 fledglings between years.

CD34+ count cell example

The dataset *AF* includes data where a new method of flow cytometry for counting CD34+ cells is compared to the readings obtained by a standard approach (Fornas et al., 2000). Both methods (coded as 1 and 3 in the dataset) were applied to a sample of 20 subjects. The aim of the study is to assess

the interchangeability between the methods so it is necessary to evaluate the degree of concordance. Hence, the ICC used here concurs with the concordance correlation coefficient estimated by variance components (Carrasco and Jover 2003).

Let's firstly consider the Poisson model. As we are facing a concordance analysis, in the *icc_counts* function the name of the method's variable (*met*) has to be provided along with the counts variable (*y*) and subject's identifier (*id*). Additionally, it is necessary to specify the concordance analysis in the *type* argument because the default is the repeatability analysis.

```
AF_P <- icc_counts(AF, y = "y", id = "id", met = "met", type = "con")
ICC(AF_P)
```

```
#>          ICC    SE ICC 95% CI LL 95% CI UL
#> [1,] 0.8472696 0.021989 0.7982025 0.8851678
```

```
VarComp(AF_P)
```

```
#>      mu    BSVAR    BMVar
#> 761.809 1.234619 0.1199439
```

The function *ICC* applied to the *iccc* object shows that the ICC estimate is 0.85 (95% confidence interval: 0.80, 0.89). Moreover, the function *VarComp* gives the parameters involved in the ICC estimator. In this case are the overall mean (*mu*), the between-subjects variance (*BSVar*) and the between-methods variability (*BMVar*) (the two last in log-scale).

Next, let's check the validity of the model by applying the function *GOF_check* to the *iccc* object.

```
set.seed(100)
AF_P.gof <- GOF_check(AF_P)
```

Figure 2a shows the plot of RQR with envelopes generated by simulation. Points on the plot stand for the RQR from the original sample. Most of points lie outside of the envelopes indicating the unsuitable fit of the model. Next plot (Figure 2b) shows the density of the RQR variances computed at the simulated samples. The RQR variance from the initial sample is 32.2 which is much larger than those from the simulations. The p-value to test overdispersion is obtained by applying the function *DispersionTest* to the *GOF*. With regards the zero inflation, no zeros were found in data so it is unnecessary to check this issue.

```
DispersionTest(AF_P.gof)
```

```
#>      S    P_value
#> 32.20049 0.00990099
```

Consequently, it is necessary to use a model able to afford the overdispersion as the Negative Binomial distribution.

```
AF_NB2 <- icc_counts(AF, y = "y", id = "id", met = "met", type = "con", fam = "nbinom2")
ICC(AF_NB2)
```

```
#>          ICC    SE ICC 95% CI LL 95% CI UL
#> [1,] 0.834794 0.0454062 0.7212048 0.9046669
```

```
VarComp(AF_NB2)
```

```
#>      mu    BSVAR    BMVar      r
#> 777.1946 1.188904 0.0809433 0.0488122
```

In this case, the ICC is quite similar to that from the Poisson model (0.83, 95% CI: 0.72, 0.90) but the confidence interval is wider. The ICC components are the same as those from the Poisson case (with different values) plus the Negative Binomial dispersion parameter (0.049). Concerning the fit of the model,

```
set.seed(100)
AF_NB2.gof <- GOF_check(AF_NB2)
```

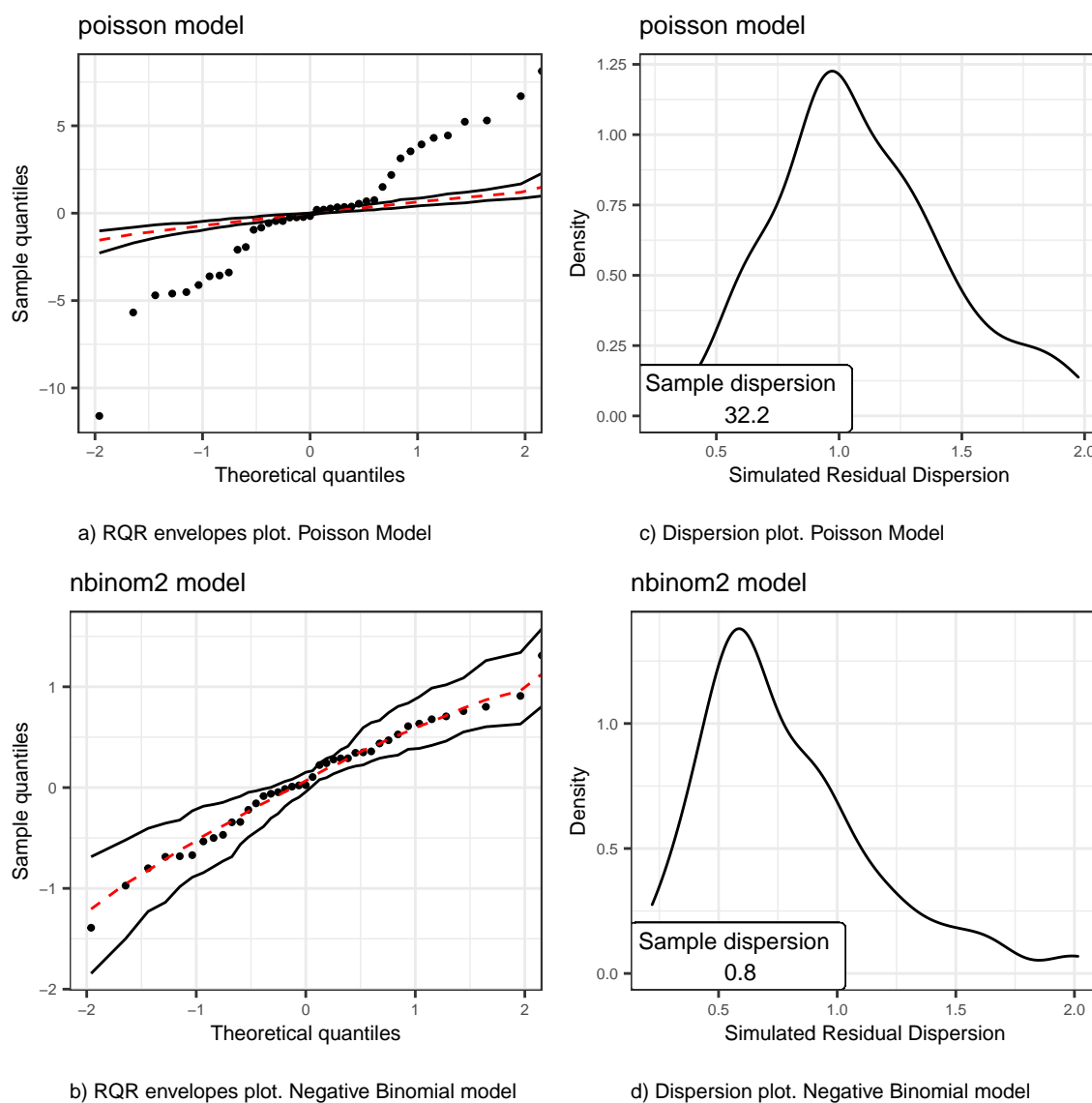


Figure 2: Goodness of fit for CD34 cell count example. The Randomized Quantile Residuals (RQR) of original data are compared to those from simulated data under the fitted model. The plots shown are RQR with envelopes, and dispersion of RQR. First row shows results for Poisson model while the plots for Negative Binomial model are on second row.

```
plot(AF_NB2.gof)
```

in Figure 2c all the RQR are within the envelopes indicating the appropriateness of the model. Additionally, the dispersion from the initial sample is within the expected values (Figure 2d). This result is also confirmed by running the dispersion test:

```
DispersionTest(AF_NB2.gof)
```

```
#>           S    P_value
#> 0.8002765 0.4059406
```

Figures 4c and 4d show the Bland-Altman plot and the Bar plot of the differences between method 1 and 2 that are generated by the following commands:

```
AF.BA <- plot_BA(AF,y="y",id="id", rm="met") # Bland-Altman plot
```

```
plot_BA(AF,y="y",id="id", type="bars") # Bar plot
```

It can be seen that for values of the mean below 700 the within-subjects differences are very close to 0. However, for larger values of the mean there is a trend in the differences in relation to the mean values.

Tick counts example

In this study, the repeatability of line transects survey method to estimate tick abundance was assessed (Kjellander et al. 2021) in the area of Grimso (Sweden). With this aim, sampling was performed by two parallel transects separated by 1m-2m where the total count of ticks was recorded. Here, the transects are the cluster variable and every pair of data from the same transect are considered as replicates. Data is stored in the package as the *Grimso* object.

As seen before the first model to consider is the Poisson model. As it is a repeatability analysis, in the *icc_counts* function we just need to provide the name of the counts variable (*Tot*) and subject's identifier (*TransectID*).

```
G_P <- icc_counts(Grimso, y = "Tot", id = "TransectID")
ICC(G_P)
```

```
#>           ICC      SE ICC 95% CI LL 95% CI UL
#> [1,] 0.3494333 0.1369518 0.0589753 0.5853431
```

```
VarComp(G_P)
```

```
#>           mu      BSVar
#> 0.2072297 1.278685
```

The function *ICC* applied to the *iccc* object shows that the ICC estimate is 0.35 (95% confidence interval: 0.06, 0.59). The function *VarComp* gives the parameters involved in the ICC estimator: the overall mean and the between-subjects variance (the latter in log-scale).

Let's apply the function *GOF_check* to the *iccc* object to check the validity of the model.

```
set.seed(100)
G_P.gof <- GOF_check(G_P)
```

```
plot(G_P.gof)
```

Figure 3a shows the plot of RQR with envelopes generated by simulation. All points on the plot lie within the envelopes indicating the fit of the model is correct. Additionally, Figure 3b shows the RQR variance from the initial sample (1.84) is compatible with the dispersion observed in the simulated samples. The overdispersion test is run by applying the function *DispersionTest* to the *GOF* object.

```
DispersionTest(G_P.gof)
```

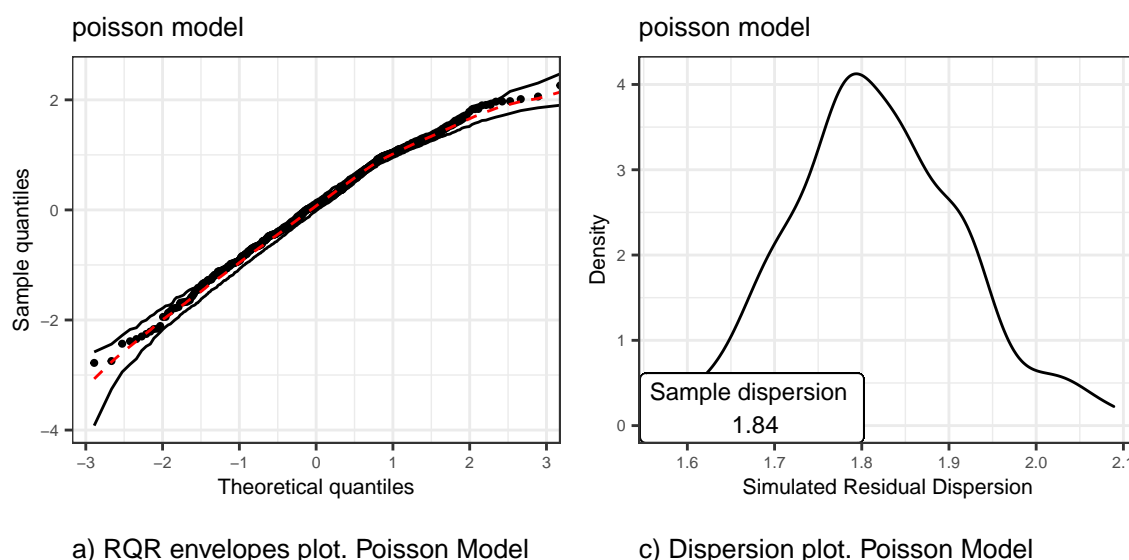


Figure 3: Goodness of fit for Tick counts example. The Randomized Quantile Residuals (RQR) of original data are compared to those from simulated data under the fitted model. The plots shown are RQR with envelopes, and dispersion of RQR for Poisson model.

```
#>      S   P_value
#> 1.83724 0.4158416
```

The p-value is 0.416 so the null hypothesis of no overdispersion is not rejected and no further models have to be fitted.

The Bland-Altman plot and the Bar plot of the within-subjects differences are shown in Figures 4e and 4f.

```
G.BA <- plot_BA(Grimso,y="Tot",id="TransectID",rm="Round") # Bland-Altman plot
```

```
plot_BA(Grimso,y="Tot",id="TransectID", type="bars") # Bar plot
```

As in the case of the sparrow fledgling paternity counts, we can observe a heteroscedastic pattern with the variability of the differences increasing with the mean. Most of the differences are 0 (75%) and 90% of the differences lie between -1 and 1.

```
quantile(G.BA$data$Diff, probs=c(0.05,0.95))
```

```
#> 5% 95%
#> -1  1
```

5 Conclusion

The statistical assessment of agreement is an issue that has received a considerable attention in recent years. It is possible to find statistical software to carry out such analysis for qualitative or continuous data (see for example Revelle (2021); Carrasco et al. (2013); D. Feng (2020)). However, there is a lack of tools when dealing with discrete data. Here, the **iccCounts** package have been introduced to assess the agreement with such type of data considering both repeatability and concordance settings. Furthermore, the **iccCounts** package provides the methodology to assess the validity of the model fitted to data.

It is important to note that no factors or predictors other than subjects and methods have been considered in the linear predictor of the GLMM. When fitting a GLMM, the inclusion of further covariates allows controlling for confounding effects. In this case, the ICC computed after controlling for confounding effects is referred to as adjusted repeatability (Nakagawa and Schielzeth 2010). Including covariates in the linear predictor make sense when the aim is to estimate the magnitude of

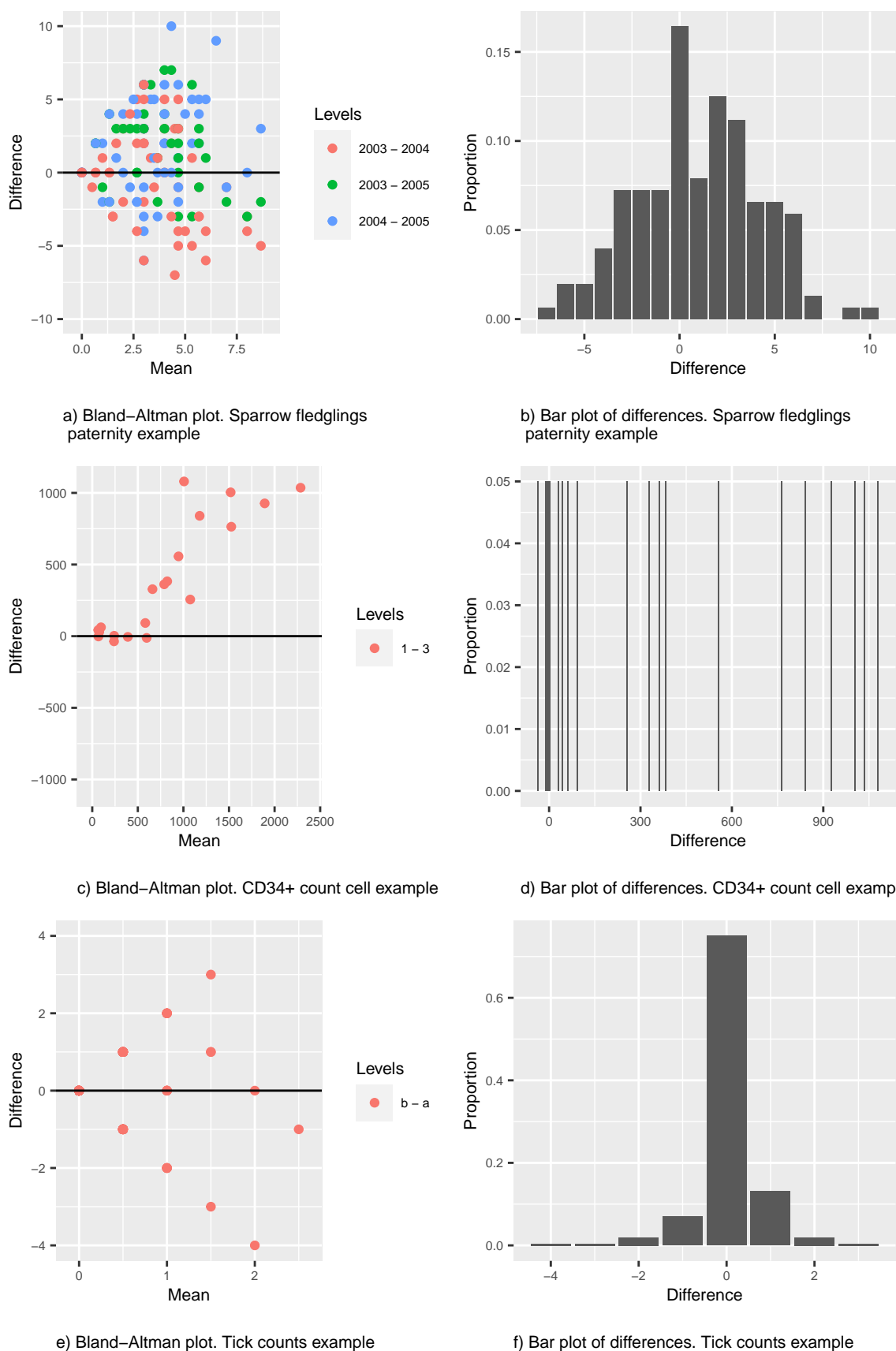


Figure 4: Bland-Altman and Bar plots. The first column shows the Bland-Altman plots where difference between pairs of data from the same subject (Y-axis) is plotted against mean of data from the same subject (X-axis). The second column contains the Bar plots of the differences between pairs of data from the same subject. The plots for Sparrow fledglings paternity example are on the first row, the CD34+ count cell example plots are on second row, and plots for Tick counts example are on third row.

an effect (difference in means, odds ratio or ratio of means, for instance) adjusted by the covariates. When estimating the ICC in a linear model, the inclusion of covariates will lead to a change in the variance components estimates but they remain as common estimates for all subjects. However, when facing the models for count data, the addition of covariates in the linear predictor leads to different ICCs because the value of the marginal mean μ will be different for every level of the covariates. For this reason, in the case of count data, it is preferable to segregate the data to estimate a different ICC according to the covariates. In this way, both the variance components and the mean will be different.

Furthermore, in the Normal model setting the ICC takes values from 0 to 1. A value of 0 means independence among the measures from the same cluster (no cluster effect) whilst a value of 1 implies perfect agreement, i.e. all data from the same subject are equal. However, it is not possible to reach a value of 1 in the counts setting. The reason for this is that some within-subject variability is unavoidable because of the relation between the variance and the mean in these models. Thus, it is not possible to observe perfect agreement with count data but the interpretation remains the same: the proportion of the total variance accounted for between-subjects variability.

6 Availability

The current stable version of the package requires R 4.0 and can be downloaded from CRAN . Furthermore, *iccCounts* depends on the following R packages: *glmmTMB* (Brooks et al. 2017); *ggplot2* (Wickham 2016); *Deriv* (Clausen and Sokol 2020); *gridExtra* (Auguie 2017); and *dplyr* (Wickham et al. 2020).

7 Acknowledgments

I would like to thank to Shinichi Nakagawa from UNSW Sydney (Australia) for kindly providing the Sparrow fledglings paternity data. I am also indebted to Pia Kjellander from Linköping University (Sweden) for sharing the Tick count data.

References

- Auguie, B. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R Package Version 2.3. <https://CRAN.R-project.org/package=gridExtra>.
- Bland, J Martin, and Douglas G Altman. 1995. "Comparing Methods of Measurement: Why Plotting Difference Against Standard Method Is Misleading." *The Lancet* 346 (8982): 1085–87.
- Bolker, Benjamin M., Mollie E. Brooks, Connie J. Clark, Shane W. Geange, John R. Poulsen, M. Henry H. Stevens, and Jada-Simone S. White. 2009. "Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution." *Trends in Ecology & Evolution* 24 (3): 127–35.
- Brooks, ME, K Kristensen, KJ van Benthem, A Magnusson, CW Berg, A Nielsen, HJ Skaug, M Maechler, and BM Bolker. 2017. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *The R Journal* 9: 299–314.
- Carrasco, Josep L. 2010. "A Generalized Concordance Correlation Coefficient Based on the Variance Components Generalized Linear Mixed Models for Overdispersed Count Data." *Biometrics* 66: 897–904.
- Carrasco, Josep L., and L Jover. 2003. "Estimating the Generalized Concordance Correlation Coefficient Through Variance Components." *Biometrics* 59: 849–58.
- . 2005. "Concordance Correlation Coefficient Applied to Discrete Data." *Statistics in Medicine* 24: 4021–34.
- Carrasco, Josep L., BR Phillips, J Puig-Martinez, TS King, and VM Chinchilli. 2013. "Estimation of the Concordance Correlation Coefficient for Repeated Measures Using SAS and R." *Computer Methods and Programs in Biomedicine* 109: 293–304.
- Choudhary, PK, and HN Nagaraja. 2017. *Measuring Agreement: Models, Methods, and Applications*. Hoboken: Wiley.
- Clausen, A, and S Sokol. 2020. *Deriv: R-Based Symbolic Differentiation*. *Deriv Package Version 4.1*. <https://CRAN.R-project.org/package=Deriv>.
- DeVet, HCW, CB Terwee, LB Mokkink, and DL Knol. 2011. *Measurement in Medicine*. Cambridge: Cambridge University Press.
- Dunn, PK, and GK Smyth. 1996. "Randomized Quantile Residuals." *Journal Computational and Graphical Statistics* 5: 236–44.
- Feng, C, L Li, and A Sadeghpour. 2020. "A Comparison of Residual Diagnosis Tools for Diagnosing Regression Models for Count Data." *BMC Medical Research Methodology* 20: 175.

- Feng, Dai. 2020. *agRee: Various Methods for Measuring Agreement*. <https://CRAN.R-project.org/package=agRee>.
- Fleiss, JL. 1986. *The Design and Analysis of Clinical Experiments*. New York: Wiley.
- Hardin, W, and JM Hilbe. 2007. *Generalized Linear Models and Extensions*. Stata Press.
- Hoef, JM Ver. 2012. "Who Invented the Delta Method?" *The American Statistician* 66: 124–27.
- Kjellander, PL, M Aronsson, UA Bergvall, Josep L. Carrasco, M Christensson, PE Lindgren, M Akesson, and P Kjellander. 2021. "Validating a Common Tick Survey Method: Cloth-Dragging and Line Transects." *Experimental and Applied Acarology* 83: 131–46.
- Nakagawa, S, and H Schielzeth. 2010. "Repeatability for Gaussian and Non-Gaussian Data: A Practical Guide for Biologists." *Biological Reviews* 85: 935–56.
- Revelle, William. 2021. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. <https://CRAN.R-project.org/package=psych>.
- Schroeder, J, T Burke, ME Mannarelli, DA Dawson, and S Nakagawa. 2012. "Maternal Effects and Heritability of Annual Productivity." *Journal of Evolutionary Biology* 25: 149–56.
- Smith, Mark W., Jun Ma, and Randall S. Stafford. 2010. "Bar Charts Enhance Bland–Altman Plots When Value Ranges Are Limited." *Journal of Clinical Epidemiology* 63 (2): 180–84.
- Stoffel, MA, S Nakagawa, and H Schielzeth. 2017. "rptR: Repeatability Estimation and Variance Decomposition by Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 8: 1639–44.
- Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wickham, H, R François, L Henry, and K Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. R Package Version 1.0.2. <https://CRAN.R-project.org/package=dplyr>.
- Wolak, ME, JF Fairbairn, and RP Paulsen. 2012. "Guidelines for Estimating Repeatability." *Methods in Ecology and Evolution* 3: 129–37.

Josep L. Carrasco

University of Barcelona

Biostatistics. Department of Basic Clinical Practice

Casanova 143, 08036, Barcelona, Spain

https://webgrec.ub.edu/webpages/personal/ang/005037_jlcarrasco.ub.edu.html

ORCID: 0000-0003-1184-0753

jlcarrasco@ub.edu