R-squared measures in Multilevel Modelling:

The undesirable property of negative R-squared values

Frist Year Paper

Research Master in Social and Behavioral Science

Student: Edoardo Costantini

Supervisor: prof. dr. M.A.L.M. van Assen

Academic year 2017/2018

Table of Contents

Measuring the explained variance by a model is a well-established way of assessing the goodness-of-fit in multiple linear regression analysis. Researchers who analyze data with a nested structure often require a measure that can be used and interpreted in the same way when applied to mixed effects models. However, the R-squared measure is not directly applicable to the multilevel context (where mixed effects models are used).

The challenge of defining and interpreting a proper measure of explained variance by a mixed model has been taken up by many researchers before. In the present paper, I analyze some of the most renowned solutions to this issue. Along with this review of the literature, a new measure, based on ANOVA decomposition of variance is proposed. All of these measures are compared based on their performance when applied to specific multilevel datasets created specifically to stress some of the strength and weakness of these measures. The focus is on establishing if the measures comply with two fundamental properties defined for measures of explained variances (Kvalseth, 1985): whether they fall in the interval 0-1 and whether their value increases as predictors are added to the models (Cameron & Windmeijer, 1996).

In the following, the concept of R-squared will be first presented in the context of multiple linear regression and it will be subsequently extended to the multilevel context (section 1). In section 2, the ANOVA decomposition-based measure will be introduced along with the review of the previously proposed measures. Subsequently, I will compare the measures performances on six fictitious example datasets (Section 3), and conclude with a discussion of the results in section 4.

1. Explained Variance in Regression Analysis (R^2)

In simple and multiple linear regression, the overall quality of the fit of a model is usually assessed through the R^2 statistics. The measure can be conceived in many different ways (for a complete overview see Kvalseth, 1985). Here, I am focusing on R^2 as a measure of explained (or modeled) variance, which is usually indirectly defined as unity minus the unexplained variance:

$$R^2 = 1 - \frac{var(Y_i - \hat{Y}_i)}{var(Y_i)} \tag{1}$$

where $var(Y_i - \hat{Y}_i)$ is the mean squared prediction error when the prediction of Y is done through a regression model, and $var(Y_i)$ is the mean squared prediction error based exclusively on the mean of the dependent variable.

To understand this statistic, it is useful to recall the concept of proportional reduction of prediction error. Intuitively, this is rooted in the idea that the error of prediction is reduced by using the prediction equation instead of the sample mean $\overline{Y}$ to predict the outcome variable, Y. More formally, the prediction error can be expressed as the difference between the observed values (Y_i) and the predicted value of Y. When no explanatory variables are used to predict the response variable, the best estimate of Y is the sample mean $\overline{Y}$ and, in such a case, we can summarize the prediction error by the squared sum of all the deviations from the mean. This is the Total Sum of Squares or $TSS = \Sigma(Y_i - \overline{Y})^2$. When explanatory variables are used in a prediction model, we talk about Residual Sum of Squares or $RSS = \Sigma(Y_i - \hat{Y})^2$.

TSS and RSS are the squared prediction errors for the null model and the model with predictors, respectively. The ratio of these two quantities is the proportional reduction in the error of prediction, a concept that is equivalent to the proportional reduction in the unexplained variance due to the inclusion of predictors in a null model. This is because TSS and RSS can be interpreted as measures of variability in the sense that they are larger the more the *y*-predictions are distant from the observed values. In this sense, TSS can be interpreted as the total variability in the outcome variable that can be explained, and RSS is the amount of this total variability that is left unexplained (not modeled) after using the explanatory variables as predictors. The ratio of these two quantities is the proportion of the total variance that is left unaccounted for by the model. If we subtract this from the total explainable variance (which is $\frac{TSS}{TSS} = 1$) then we have the amount of explained variance by the model:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\Sigma(Y_i - \hat{Y})^2}{\Sigma(Y_i - \bar{Y})^2} = 1 - \frac{var(Y_i - \hat{Y})}{var(Y_i)} \quad (2)$$

Two properties of the defined R^2 measures are especially important: (1) R^2 will always fall between 0 and 1; (2) adding predictors to the model cannot decrease the value of R^2.

For proving the first property, the difference between the predicted value $\hat{Y}_i$ and the observed Y_i can be expressed as:

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) - (\hat{Y}_i - \bar{Y})$$

by taking the square of both sides and summing across all observations *i*, we obtain:

$$\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2 = \Sigma_{i=1}^{n}\left[\left(Y_i - \hat{Y}_i\right) - \left(\hat{Y}_i - \bar{Y}\right)\right]^2$$

which can be rewritten by algebraic manipulation[1] (e.g. Draper & Smith, 1981) as

$$\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2 = \Sigma_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \Sigma_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

which makes obvious that $\Sigma(Y_i - \hat{Y}_i)^2 \leq \Sigma(Y_i - \bar{Y})^2$, and therefore $R^2 \geq 0$.

As for the second property, it is also clear from equation 1 that R-squared can decrease from a model to another only if the RSS increases from a first to a second model. However, adding predictors to a linear regression model will never increase the prediction error since it is not possible to explain less variation in *Y* by adding a predictor to a linear model (Agresti & Finlay, 2009).

2. The challenges of R^2 Measures for Mixed Models

The first challenge in the application of the R^2 statistic to hierarchical linear modeling is the presence of multiple variance components (e.g., within-group variance, between-group variance,

---

[1] The cross-product of resulting from $\left[(Y_i - \bar{Y}) - (\hat{Y}_i - \bar{Y})\right]^2$ is equal to 0: $CPT = 2\Sigma\left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \bar{Y}\right) = 0$

Since $Y_i - \hat{Y}_i = Y_i - \bar{Y} - b_1\left(X_i - \bar{X}\right)$

And $\hat{Y}_i - \bar{Y} = b_1\left(X_i - \bar{X}\right)$

Then
$$CPT = 2\sum b_1\left(X_i - \bar{X}\right)\left[Y_i - \bar{Y} - b_1\left(X_i - \bar{X}\right)\right] =$$
$$= 2\sum\left[b_1\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) - b_1^2\left(X_i - \bar{X}\right)^2\right] =$$
$$= 2b_1(\sigma_{XY} - b_1\sigma_X^2) = 0$$

since b_1 is the ratio of the covariance between X and Y (σ_{XY}) and the variance in X (σ_X^2), $b_1 = \frac{\sigma_{XY}}{\sigma_X^2}$.

slope variance). A first decision must be made regarding what level is the explained variance measure referring to, $var(Y_{ij})$ or $var(\bar{Y}_j)$. A second, related, problem is that the presence of random slopes in the model makes the definition of explained variance even more complex; when random slopes are included in the model, the explained and unexplained variances depend on the values of the explanatory variable (heteroscedasticity).

To understand the challenges that R-squared measures face in multilevel analysis, it is helpful to consider, as examples, two different mixed models: one random intercept and one random slope model. Both models will have one dependent variable *Y*, one level-1 predictor X_{ij}, and one level-2 predictor W_j.

We begin with the random intercept model by presenting the relationship between the outcome variable *Y* and its predictors within any group *j* (level-1 model).

$$Y_{ij} = \beta_{0j} + \beta_1 X_{1ij} + e_{ij} \quad (3)$$

The intercept β_{0j} is the expected *Y* score for all individuals in a group *j*, β_1 is the slope, common to all groups, which indicates the average expected change in *Y* associated with a unit change in *X*, the level-1 predictor. Finally, e_{ij} is the unique effect of person *i* in group *j*, also known as error variance, which is assumed to be standard normally distributed, $e_{ij} \sim N(0, \sigma^2)$.

The subscript *j*, assigned to the intercept, indicates that its value is allowed to change between groups. Its specific value depends on a constant, common to all groups (which might be interpreted as the true grand mean), and a group specific effect u_{0j}, which is also assumed to be standard normally distributed $u_{0j} \sim N(0, \tau_0^2)$[2]. In the model we are describing here, there is also a group-level variable W_j which is going to be part of the specific model of the intercept. All this is expressed in the following level-2 equations:

[2] The 0 as subscript in the notation τ_0^2 serves to distinguish the variance of the random intercepts, from the variance of the random slopes τ_1^2 which will be introduced in the sequel.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_1 = \gamma_{10} \tag{4}$$

Equations 3 and the set of equations 4 are often referred to as level 1 and level 2 equations of the multilevel model, respectively. By plugging the latter into the former we obtain the actual random intercept mixed model (or random ANCOVA model) with a predictor at each level:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + u_{0j} + e_{ij} \tag{5}$$

This final model has five parameters in total, three for the fixed effects ($\gamma_{00},\ \gamma_{10},\ \gamma_{01}$), and two for the random effects (σ^2, τ_0^2).

The second model that is useful to describe is the random slope mixed model. The individual level relationship remains mostly unchanged, except for the regression coefficient of the predictor which now gains a subscript *j*. As for the intercept in the previous case, this subscript indicates that the slope has now a group-specific value.

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + e_{ij} \tag{6}$$

The part of the set of level-2 equations that models the intercept variation remains unchanged, while in the equation corresponding to the regression coefficient, a new group specific effect comes into play, namely u_{1j}:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{7}$$

where u_{*j} belongs to a multivariate standard normal distribution, that is $u_{*j} \sim N(0, \sum)$. Since random slopes and intercepts are usually correlated, three parameters must be defined as part of the variance-covariance matrix $\sum$. These are the random intercept variance τ_0^2, the random slope variance τ_1^2, and the covariance between the two τ_{01}:

$$\Sigma = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}$$

The resulting mixed model will hence be the following:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + u_{1j}X_{ij} + \gamma_{01}W_j + e_{ij} + u_{0j} \tag{8}$$

which has a total of seven parameters: $\gamma_{00},\ \gamma_{10},\ \gamma_{01},\ \sigma^2,\ \tau_0^2,\ \tau_1^2,\ \tau_{01}$. It is important to note that as a consequence of the introduction of random slopes, the error variance in *Y* now depends on the value of *X*, what is known as heteroscedasticity (Snijders & Bosker, 1999, Chapter 8). In other words, in certain groups, the value of *X* might not matter at all for the prediction of the *Y* outcome, while in other groups it might be quite influential.

2.1 Understanding variance: variance components in nested data

Equation 5 and 8 are modeling the response variable *Y*. The observed variance in *Y* can be decomposed into the empty (or null) model level-1 and level-2 variances, σ_0^2 and τ_{00}^2 respectively[3]. The null model, which is needed for the definition and calculation of the explained variance, is a model stripped of all its predictors and in the multilevel context it often takes the following form:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \tag{9}$$

The total variance in the observed values of Y_{ij} is the sum of these two variances:

$$var(Y_{ij}) = \sigma_0^2 + \tau_{00}^2 \tag{10}$$

The estimation of these two parameters in the null model allows to define the intraclass correlation coefficient (ρ), which expresses the percentage of the total explainable variance in Y_{ij} that is accounted by the group membership:

$$\rho = \frac{\tau_{00}^2}{\tau_{00}^2 + \sigma_0^2} \tag{11}$$

Once we add predictors to the null model, the meaning of ICC changes. If we were to estimate the variance components of the model in equation 5 for example, we would have to interpret ρ (now =

[3] The second 0 in the notation of τ_{00}^2 allows to distinguish the group level variance in the null model from the unexplained level-2 variance of a full model τ_0^2. It has the same meaning of the 0 I use to distinguish the error variance in the null model, σ_0^2, from the error variance in the full model σ^2.

$\frac{\tau_0^2}{\tau_0^2+\sigma^2}$) as the proportion of variance in *Y* accounted for by the group membership, controlling for the predictors included in the model. In more intuitive terms, it would be the proportion of the variance in Y_{ij}, unexplained by the model, that is attributable to the clustering of first level units in second level units, keeping the covariates constant.

When random slopes are added to the model, the concept of ICC has no longer any fixed meaning because adding the random slopes leads to heteroscedasticity and therefore the error variance depends on the values of individual level predictors. As mentioned above, in the random slopes model, the level-2 random effects dispersion is defined not only by their variances, but also by their covariances: τ_{01}.

Thus far, only the variance in Y_{ij} has been discussed. However, predictions can be made also at the second (or higher) level when predicting the value of the group mean $\bar{Y}_j$, for a randomly selected level-2 unit j. According to Snijders and Bosker (1994), in the null model, the (unexplained) variance at this level is then expressed by:

$$var(\bar{Y}_j) = \tau_{00}^2 + \frac{\sigma_0^2}{n_j} \quad (12)$$

Which is the sum of the true variance between the groups and the expected sampling error variance. This latter quantity is the quotient of the within-group variance and the harmonic mean of the group means sizes. When the group are small, the expected sampling error variance has a substantial effect in determining the expected observed variance between groups; on the contrary, when groups are large, the observed between groups variance, $var(\bar{Y}_j)$, is much closer to the true between group variance, τ_{00}^2.

When the random intercept model with predictors is considered, the level-2 unexplained variance is expressed as:

$$var(\bar{Y}_j|X) = \tau_0^2 + \frac{\sigma^2}{n_j} \quad (13)$$

Table 1 reports a summary of all the variances that have been described and their variance components when applicable.

Table 1 – *Variances and variance components*

	Description	Variance components
$var(Y_{ij})$	Variance in the observed values of the response variable in the empty or null model	$\sigma_0^2 + \tau_{00}^2$
$var(Y_{ij} - Y'_{ij})$	Unexplained variance in the observed values of the response variable in the model with predictors	$\sigma^2 + \tau_0^2$
$var(\bar{Y}_j)$	Variance in the average of the observed response variable in group *j* in the null model	$\tau_{00}^2 + \frac{\sigma_0^2}{n_j}$
$var(\bar{Y}_j - \bar{Y}'_j) = var(\bar{Y}_j \vert X)$	Unexplained variance in the average of the observed response variable in group *j* in the model with predictors	$\tau_0^2 + \frac{\sigma^2}{n_j}$
ICC	Intraclass correlation	$\frac{\tau_{00}^2}{\tau_{00}^2 + \sigma_0^2}$
$var(e_{ij})$	Variance in the unique effect of individual *i* in group *j*	σ^2
$var(u_{0j})$	Variance in the unique effect of group *j* on the mean response variable, holding all predictors constant	τ_0^2
$var(u_{1j})$	Variance in the unique effect of group *j* on the slope of the level-1 predictor	τ_1^2
$cov(u_{0j}, u_{1j})$	Covariance between level-1 intercepts and slopes	τ_{01}

2.2 Level-specific measures of modelled variance

In the following, first, two wildly used measures of level-specific R^2 will be presented and then their characteristics will be discussed in detail. These first statistics allow to distinguish the explained variance at different levels. Subsequently, two measures that assess the overall explained variance will be discussed. Finally, a comprehensive alternative measure based on ANOVA decomposition of variance will be introduced.

2.2.1 Bryk and Raudenbush (1992)

Bryk and Raudenbush (1992) have proposed one of the first measures of explained variance as a relative measure. Their measure expresses the reduction in unexplained variance when predictors are added to a null model.

As explained above, in multilevel analysis, fitting the null model provides an estimate of the proportions of variance in Y_{ij} that are due to the group or the individual level. When predictors are added, and a random intercept mixed model is estimated, part of the total variance in the outcome variable is now modelled by those predictors, and part of it is still unaccounted for. An estimate of this unaccounted part is provided by the estimation of the variances of the random effects (σ^2, and τ_0^2). This provides an opportunity to define a proportion of reduction in (unaccounted) variance, at both levels, or in other words, a measure of modelled variance. The approach illustrated by Bryk and Raudenbush (1992) consists of comparing the residual error variances, *within each level*, of a random intercept model with predictors, with the total variance partitions estimates of the random intercept-only model (null model). Therefore, the estimated amount of explained variance at the second level is:

$$R^2_{2B\&R} = \frac{\tau^2_{00} - \tau^2_0}{\tau^2_{00}} = 1 - \frac{\tau^2_0}{\tau^2_{00}} \tag{14}$$

Likewise, at the first level:

$$R^2_{1B\&R} = \frac{\sigma^2_0 - \sigma^2}{\sigma^2_0} = 1 - \frac{\sigma^2}{\sigma^2_0} \tag{15}$$

where the footer *0* refers to the null model and the variances without it regard the mixed model with predictors.

The sum of the proportions resulting from equation 14 and 15 can be interpreted as the total amount of variance explained by the model:

$$R^2_{B\&R} = R^2_{1B\&R} + R^2_{2B\&R} \tag{16}$$

The code for measures implementation in R of these and all the following approaches are presented in appendix 1 with detailed comments.

2.2.2 Snijders and Bosker (1994)

A second widely appreciated approach has been proposed by Snijders and Bosker (1994). Their starting point is an attentive decomposition of the variance of the prediction errors at the individual, as well as at the group level.

Starting from level-1, they point out that, in the model with prediction, the variance of the prediction error (MSE) of the individual values on the response variable Y_{ij} is the sum of the residual variances at both levels estimated in the full model:

$$var\left(Y_{ij} - \hat{Y}_{ij}\right) = \sigma^2 + \tau_0^2 \tag{17}$$

By considering the variance of the prediction error for both the null model and the model with predictors, and applying the standard definition of R^2 given in equation 1, they straightforwardly derive the following level-1 explained variance measure:

$$R^2_{1S\&B} = 1 - \frac{var\left(Y_{ij} - \hat{Y}_{ij}\right)}{var\left(Y_{ij}\right)} = 1 - \frac{\left(\hat{\sigma}^2 + \hat{\tau}_0^2\right)}{\left(\hat{\sigma}_0^2 + \hat{\tau}_{00}^2\right)} \tag{18}$$

(notation as above).

There is a double interpretation of $R^2_{1S\&B}$: its value can be referred to as the contribution to of the predictors to the explained variance at the first level, or as the proportional reduction in the value of the total unexplained variance $(\sigma^2 + \tau_0^2)$. The latter interpretation allows to consider this measure as a measure of total explained variance.

For what concerns the second level, the reduction in the mean square prediction error of the group means is considered. In the null model, when no predictors are present, the best predictor of $\bar{Y}_j$ is its expectation, and the associated MSE is $var\left(\bar{Y}_j\right)$, expressed as in equation 12. In the random intercept model with predictors, the best predictor of $\bar{Y}_j$ is the outcome of $\bar{X}_j\beta$ and the associated prediction error variance is:

$$var\left(\bar{Y}_j - \bar{X}_j\beta\right) = var\left(\bar{Y}_j - \bar{Y'}_j\right) = var\left(\bar{Y}_j | \mathrm{X}_j\right) = \frac{\sigma^2}{n_j} + \tau_0^2 \qquad \text{(19, rep. 13)}$$

Hence, by applying equation 1 and substituting the relevant estimates of the variance of the prediction error for the second level, Snijders and Bosker define the estimated level-2 explained variance as the proportional reduction in the means square prediction error of the group means:

$$R^2_{2S\&B} = 1 - \frac{var(\bar{Y}_j - \bar{Y}'_j)}{var(\bar{Y}_j)} = 1 - \frac{\left(\frac{\sigma^2}{n} + \tau_0^2\right)}{\left(\frac{\sigma_0^2}{n} + \tau_{00}^2\right)} \tag{20}$$

It is important to consider the differences between the two approaches just reviewed. Bryk and Raudenbush's (1992) measures are focused on the variances of the random effects at the two levels. Equations 14 and 15 take into account only the variance of the random effects of the group means, and the error variance (respectively) to estimate the level specific explained variances.

The measures Snijders and Bosker (1994) proposed uses the variance components of the individual scores (which means both error variance and group level randomness, compare equation 15 and 18) to compute the measure of first-level explained variance. Furthermore, the level-2 modelled variance measure is based on a different conceptualization of the unexplained between-group variability. While Bryk and Raudenbush consider τ_0^2, the variance of the random effects (i.e. intercept variance), to fully account for the unexplained between group-variability of Y, Snijders and Bosker define the group level unexplained variability as the variance in the group means of Y, conditionally on the predictors, $var(\bar{Y}_j|\mathrm{X}_j)$. This allows the group-level unexplained variance used in their formula to be broken down into two components: the random effect variance (τ_0^2) and the (weighted) error variance $\left(\frac{\sigma^2}{n_j}\right)$, see equation 14 and 20.

Snijders and Bosker (1994, 1999) discuss the extension of their measures to the random slope model. They present the detailed computational processes to obtain the variance estimates for models with both intercept and random slopes. However, by computing the R-squared measure for the intercept and random slope model in their proposed way, and the R-squared measures for the same model without the random slopes, with equation 18 and 20, they show that the results do not differ meaningfully. Hence, they conclude that the best way to compute the explained variance by a random slope model is to estimate the same model *without* specifying the randomness of the slopes and computing R^2 with equations 18 and 20.

2.2.3 Variance components-based measures undesirable characteristics

Despite their intuitive interpretation and widespread use, both these measures have two characteristics that we believe are undesirable. Recall that two of the properties of an explained variance measure are that its values should fall between 0 and 1 (Kvalseth, 1985), and that by adding a predictor to the model, it should always increase (Cameron & Windmeijer, 1996).

The problem of negative R^2 values might arise, for example, with the *B&R* measure when adding a level-1 predictor with no variation at the higher level. As Snijders and Bosker (1994) illustrated, in such a case, the $R^2_{2B\&R}$ may result in negative values. The reason lies in that it is not unusual to observe an increase in the level 2 unexplained variance when adding variables to a model. This can be understood in intuitive terms by looking at equation 19. If a within-group deviation variable[4] is added to a model, this will decrease the error variance σ^2. However, adding such a variable cannot affect the between-group unexplained variance. Therefore, the conditional variance of the group means of Y ($var(\bar{Y}_j|\mathrm{X}_j)$) will remain unchanged, σ^2 will reduce, and τ_0^2 must increase to compensate. When applying the B&R level-2 explained variance measure, τ^2 is compared between the null to the full model (see equation 14) and if τ^2 increases by adding predictors, the result will be a negative R-squared measure (Sniders & Bosker, 1994, pp. 346-350).

Snijders and Bosker's (1994) measures do not suffer from the drawback of negative R-squared measure because of an increased τ^2 when adding a within-group deviation variable. The reason is that the parameter used for comparison, of the null and full model, when computing the R-squared measure, is not directly the intercept variance, τ^2, as in equation 14, but the second level variance in

[4] A group mean centered variable, by definition, has zero between-group variance.

Y controlling for the predictors, $var(\bar{Y}_j|\mathrm{X}_j)$, as in equation 20. As explained in section 2.2.2, this is a crucial difference between the measures proposed by B&S and S&B. Using in the computation of R-squared the parameter $var(\bar{Y}_j|\mathrm{X}_j)$, as computed in equation 19, allows to keep into account the decrease in the estimated σ^2 and the increase in the estimated τ^2, that occurs when adding a group-mean centered level-1 predictor, hence, avoiding the increase in level-2 residual variance from the null to the full model. However, the S&B measures are not exempt from the undesirable feature of negative level-2 R-squared values as our following fictitious example shows.

The example consists of 60 cases divided into 10 balanced groups. Two variables are present in the dataset: Y and X, which respectively assume values between 0 and 4, and 1 and 30. The first group has a mean value of X equal to 2, for the second group it is equal to 5, and each for each group, the mean is incremented by 3 units (8, 11, 14, etc.) until 29, the mean of the 10th group. In each group the relationship between the predictor X and the outcome Y can be defined by the regression equation: $Y = b_0 + b_1X + e$, The grand mean $\bar{Y}_{.}$ and the group averages $\bar{Y}_j$ are all equal to 2. In other words, there is no variation of the group means around the grand mean. See Figure 1 for a scatter plot of the data set of the example.

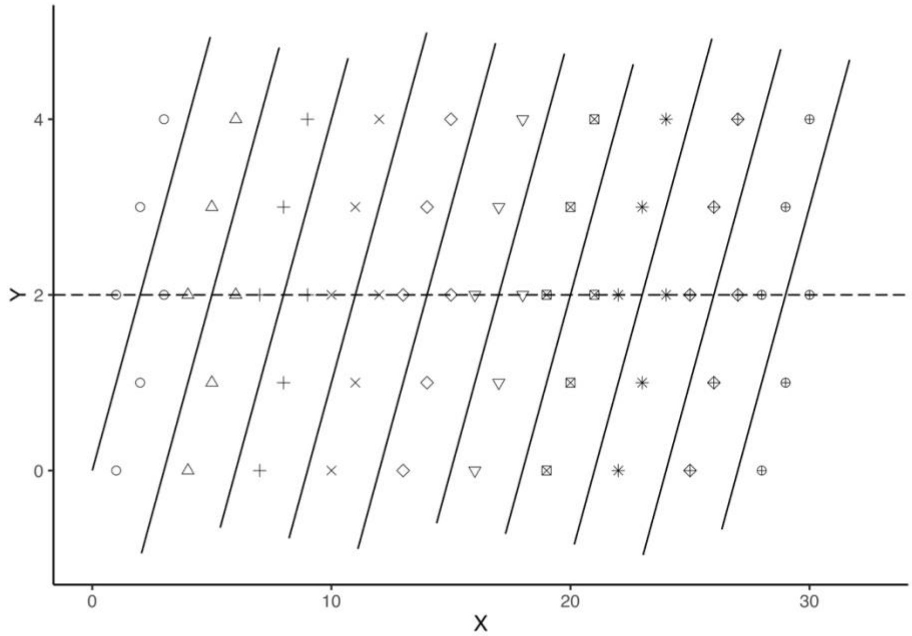

Figure 1 – *Fictitious dataset illustrating a scenario in which both B&R and S&B measures of level-2 explained variance provide uninterpretable results.*

If we consider the random intercept *only* (null) model for the fictitious dataset the total variance in the observed Y_{ij} is due only to the individuals' unique effects (e_{ij}) since u_{0j} is equal to 0 in each group. By adding to the model the group means as a second level variable the prediction is not improved in any way and the variance of Y conditional on X stays 0. However, when the within-group deviation score variable is added to the model as well, the between group variance of Y conditional on X (unexplained level-2 variance) increases sharply. The reason is that, as it is visible from Figure 1, controlling for predictor X, groups differ substantially in their "average" or predicted value of Y; adding a predictor to the random intercept only model (or null model) has "created" a large amount of unexplained variance in Y_{ij}, at the second level. This will result in negative values of R^2 for both $R^2_{2B\&R}$ and $R^2_{2S\&B}$, which can be computed using the estimates of the variances of the random effects reported in Table 2.

Table 2 – *Variance components estimated for null (first row) and full (second row) mixed model applied to the arbitrary data just described.*

Model	$\hat{\sigma}^2$	$\hat{\tau}_{00}^2$
$Y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$	1.667	0
$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \mu_{0j} + e_{ij}$	1.004	81.791

Applying equations 15 and 14:

$$R_{1B\&R}^2 = 1 - \frac{1.004}{1.667} = .398$$

$$R_{2B\&R}^2 = 1 - \frac{81.791}{0}$$

While the $R_{1B\&R}^2$ can be computed and is interpretable as the proportion of explained variance at the first level by the model with the predictor, compared to the null model, the $R_{2B\&R}^2$ measure has no mathematical meaning. However, if there had been any variance at the second level to begin with, R_2^2 would be a large negative number. For what concerns the S&B measure, results are even more undesirable: $R_{1S\&B}^2 = -48.667$ and $R_{2S\&B}^2 = -2943.490$.

The example highlights how the undesirable characteristics of both $R_{*B\&R}^2$ and $R_{*S\&B}^2$ are more deeply rooted in the conception of the modelled variance as a measure of reduction in variance components. This feature makes the performance of both measures strongly dependent on the definition of the model, since, just by adding a level-1 predictor, the variance components at the higher-level change interpretation. The reason is that by including a level-1 predictor, the interpretation of τ^2 changes: it is the variance, controlled for all predictors. As the predictors change, so does the interpretation of τ^2.

Snijders and Bosker (1994) would argue that the negative measures are evidences of a misspecified model. In particular, they would advise to use a group mean centered transformation of the predictor X, instead of the raw variable. However, as will be shown in the remaining of this article, their measures can result in negative values even when the model is correctly specified.

2.3 Overall explained variance

Thus far, I have presented measures of modelled variance by the models within the different levels. However, the overall variance explained by a model can also be of interest. As I mentioned before, both the measures proposed by Bryk and Raudenbush (1992), and the ones proposed by Snijders and Bosker (1994) have a way of accounting for it. However, there are also other overall measures that can be considered. In the remaining of this section, two such measures are illustrated, and their performance will be assessed in the empirical section.

2.3.1 Nakagawa and Schielzeth (2013)

Nakagawa and Schielzeth (2013) defined a measure of overall explained variance that allowed for the distinction of the contribution to the explanation of the total variance in Y provided by the fixed and the random effects specified in a model. However, this measure is restricted to random intercept models.

Nakagawa and Schielzeth (2013) claim that R^2 measures can be grouped in two families: *marginal* R^2, which consider the variance explained by fixed effects, and *conditional* R^2, which describe variance explained by both fixed and random effects. The terms *marginal* and *conditional* come from the literature that extends linear mixed models to account non-linear relationships (e.g. Vonesh, Chinchilli, and Pu, 1996). The same terminology will be used in order to maintain continuity with the literature, but the focus here is put only on how the conditional measure is an interesting alternative to express the overall explained variance by a model. Nakagawa and Schielzeth (2013) refer to the *marginal* *R*-squared measure as the proportion of the total variance modelled by the *fixed* effects:

$$R_m^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sum_{l=1}^{u} \sigma_l^2 + \sigma_\varepsilon^2} \tag{21}$$

$$\sigma_f^2 = var\left(\sum_{p=1}^{P} \beta_p X_{pij}\right)$$

In the numerator, σ_f^2 is the variance of the predictions based on the fixed effects (*p* is the p^{th} predictor), the modelled variance. In the denominator, there is the total variance in the outcome variable given by the sum of the modelled variance, σ_f^2, and the not-modelled variance, the sum of σ_ε^2, the error variance, and all the σ_l^2, the variances of the *l*th of *u* random effects[5]. To understand what σ_l^2 represents, consider a multilevel dataset where observations are nested in *u* higher-level units, where *u* is the number of nesting levels. If there are only two levels (e.g. students, and classes), there is only one higher level nesting unit (the classroom). Therefore, the term $\sum_{l=1}^{u} \sigma_l^2$ would just be equal to σ_{CLASS}^2, the group level variance (previously referred to as τ_0^2). If the dataset included a third level, for example the school grouping, then there would be two higher level nesting units (classroom and school) and therefore $\sum_{l=1}^{u} \sigma_l^2$ would be equal to $\sigma_{CLASS}^2 + \sigma_{SCHOOL}^2$. Since in the present paper we are limiting the scope to two-level data structures, R_m^2 of equation 21 simplifies to:

$$R_m^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_l^2 + \sigma_\varepsilon^2} \quad (22)$$

(henceforth referred to as $R_{N\&S(m)}^2$).

This can simply be interpreted as the portion of the total variance, expressed as the sum of the variance in the predicted scores plus the residual variances at the first and second level, that is modelled by the variance of the predicted scores.

The *conditional R*-squared measure is defined as the proportion of the total variance modelled by the *fixed and* the *random* effects together. It is expressed as:

$$R_c^2 = \frac{\sigma_f^2 + \sum_{l=1}^{u} \sigma_l^2}{\sigma_f^2 + \sum_{l=1}^{u} \sigma_l^2 + \sigma_\varepsilon^2} \quad (23)$$

which in the case of datasets with only two levels simplifies to:

---

[5] It is important to note here that by saying *l*th random effect Nakagawa and Schielzeth (2013) are not distinguishing, for example, between a σ_1^2 variance of the random intercept, and a σ_2^2 variance of the random slopes, but they are distinguishing between a σ_1^2 variance of the random intercept due to the nesting of units in one group-level, and σ_2^2 variance of the random intercept due to the nesting of the groups of units into higher level groups.

$$R_c^2 = \frac{\sigma_f^2+\sigma_l^2}{\sigma_f^2+\sigma_l^2+\sigma_\varepsilon^2} \tag{24}$$

(henceforth referred to as $R^2_{N\&S(c)}$).

This measure is a generalization of the one proposed by Snijders and Bosker (1994) to the generalized linear mixed effects models (GLMMs) that also allows to estimate both the overall explained variance by a model and the variance at each specific level (see Nakagawa and Schielzeth, 2013, for details on the level-specific extension). It can be interpreted as the variance explained by the entire model and it does not require the specification of a null model as reference. However, it is important to notice that this measure counts the random effects as explained variances. In practice, computing both $R^2_{N\&S(m)}$ and $R^2_{N\&S(c)}$ and subtracting the first from the second can give a measure of how much of the total variance is due to the random effects.

One of the downsides of this measure is that it is not straightforwardly applicable to the context of random intercept models. An interested reader should consult Johnson (2014) who proposed an interesting approach in this direction.

2.3.2 Xu (2003)

Xu (2003) reviewed different approaches to the issue of multilevel modelled variance and proposed one based on the estimation of the posterior means by use of the Empirical Bayes estimator. The measure uses the residuals under an empty and a mixed model in analogy with ordinary linear regression as follows. The residuals are computed under both models as the differences between observed and predicted values of Y, where the predicted values are obtained after estimating the random effects by an empirical Bayes estimator (EB). In particular, the predicted values by the mixed model are expressed as:

$$\hat{\boldsymbol{y}}_{ij} = \hat{\boldsymbol{\beta}}' X_{ij} + \hat{\boldsymbol{u}}_{0j}^{EB} W_{ij}$$

with standard matrix algebra notation using the apostrophe to express the transpose of a vector. X_{ij} is the vector of the predictors with fixed effects, and W_{ij} is the vector of the covariates with random effects as estimated using empirical Bayes. In a model with only one random effect (random intercept), i.e. with W_{ij} equal to 1, the formula above simplifies to:

$$\hat{\boldsymbol{y}}_{ij} = \hat{\boldsymbol{\beta}}' X_{ij} + \hat{\boldsymbol{u}}_{0j}^{EB}$$

Similarly, the prediction done with the empty model takes the form:

$$\hat{\boldsymbol{y}}_{ij} = \hat{\gamma}_{00} + \hat{\boldsymbol{u}}_{0j}^{EB}$$

$\hat{\boldsymbol{u}}_{0j}^{EB}$ is the vector of empirical Bayes estimated group random-effects, which can be computed for the null model as:

$$\hat{\boldsymbol{u}}_{0j}^{EB} = \boldsymbol{\lambda}_j \hat{\boldsymbol{\beta}}_{0j} + (1 - \boldsymbol{\lambda}_j)\hat{\gamma}_{00}$$

where $\hat{\beta}_{0j} = \bar{Y}_{.j}$ denotes the group mean, and $\boldsymbol{\lambda}_j$ is the vector the reliabilities of all groups j ($\lambda_j = \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n_j}$).

The R^2 measure proposed by Xu (2003) is then computed as:

$$R_{Xu}^2 = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{ni}\left(y_{ij} - \left(\hat{\boldsymbol{\beta}}' X_{ij} + \hat{\boldsymbol{u}}_{0j}^{EB}\right)\right)}{\sum_{i=1}^{n}\sum_{j=1}^{ni}\left(y_{ij} - \left(\hat{\gamma}_{00} + \hat{\boldsymbol{u}}_{0j}^{EB}\right)\right)} = 1 - \frac{RSS}{RSS_0} \quad (25)$$

Which is the proportion of reduction in the variance of the prediction error that occurs when adding to the null model predictors for which fixed and random effects are specified. In other words, R_{Xu}^2 measures the proportion of the variation in the outcome variable that is explained by the explanatory variables included in the model.

This measure is a measure of *overall* explained variance in the sense that it does not distinguish between a level-1 and level-2 variances to be explained, but it concerns itself exclusively with explaining the variance of individual responses as sole indicator of the fit of a model.

2.4. ANOVA variance decomposition measure

Given the conceptual similarity between AN(C)OVA and hierarchical linear modelling, the issue of explained variance in the latter can be addressed by starting with the decomposition of variance of the former. According to the usual decomposition of the total variability around the overall mean (TSS) as defined in traditional ANOVA, we can rewrite TSS as the sum of the between-groups sum of squares (SS_B), and the within-groups sum of squares (SS_W):

$$\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij}-\bar{y}\right)^2=\sum_{j=1}^{J}n_j\left(\bar{y}_j-\bar{y}\right)^2+\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij}-\bar{y}_j\right)^2$$

where y_{ij} are the observed y values, $\bar{y}$ is the observed grand mean, $\bar{y}_j$ are the observed group means of y.

By summing and subtracting to both the within- and between-group squared differences the predicted values at the corresponding level (y'_{ij}, the predicted response variable, and $\bar{y}'_j$, the average of the predicted values in each group), SS_B and SS_W can be rewritten in terms of predicted and unpredicted squared differences:

$$SS_B=\sum_{j=1}^{J}n_j\left(\bar{y}_j-\bar{y}\right)^2=\sum_{j=1}^{J}n_j\left[\left(\bar{y}_j-\bar{y}'_j\right)^2+2\left(\bar{y}_j-\bar{y}'_j\right)\left(\bar{y}'_j-\bar{y}\right)+\left(\bar{y}'_j-\bar{y}\right)^2\right]$$

$$SS_W=\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij}-\bar{y}_j\right)^2=\sum_{j=1}^{J}\sum_{i=1}^{I}\left[\left(y_{ij}-y'_{ij}\right)^2+2\left(y_{ij}-y'_{ij}\right)\left(y'_{ij}-\bar{y}_j\right)+\left(y'_{ij}-\bar{y}_j\right)^2\right]$$

In these rewritten forms of SS_B and SS_W, the terms to the left ($\sum_{j=1}^{J}n_j\left(\bar{y}_j-\bar{y}'_j\right)^2$ and $\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij}-y'_{ij}\right)^2$) represent the unpredicted parts of the variance, whereas the terms to the right ($\sum_{j=1}^{J}n_j\left(\bar{y}'_j-\bar{y}\right)^2$ and $\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y'_{ij}-\bar{y}_j\right)^2$) represent the parts predicted by the model. These sums of squares can be used to evaluate the predicted variance in percentage over the total variance. SS_B and SS_W can be considered as the total amount of variance at each level.

Therefore, we can compute the level 1 and level 2 explained variance measures as:

$$R^2_{1ANV}=\frac{\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y'_{ij}-\bar{y}_j\right)^2}{\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij}-\bar{y}_j\right)^2} \qquad (26)$$

$$R^2_{2ANV}=\frac{\sum_{j=1}^{J}n_j\left(\bar{y}'_j-\bar{y}\right)^2}{\sum_{j=1}^{J}n_j\left(\bar{y}_j-\bar{y}\right)^2} \qquad (27)$$

In the hierarchical linear model context, the predicted value of the outcome variable (y'_{ij}) at level 1 can be obtained using any estimation method. The level-2 predicted values, the group means $\bar{y}'_j$, are obtained by taking the average group values of the level-1 predicted values y'_{ij}.

Finally, we can also describe the total variance explained by the model as:

$$R^2_{ANV} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y'_{ij}-\bar{y}\right)^2}{\sum_{j=1}^{J}\sum_{i=1}^{I}\left(y_{ij}-\bar{y}\right)^2} \tag{28}$$

When interpreting this measure, it is important to stress that it is expressing the amount of explained variance (within, between, or total) by the model compared to a null model without any predictor *and* without any random effects. This is a crucial difference compared to the other measures discussed, where the null model is usually defined to have no predictors but to have random effects (i.e. random-intercept *only* model). This means that this measure expresses how much of the variance is model by the addition of the predictors *and* the specification of the random effects.

A defining characteristic of the ANOVA decomposition measure, which is here deemed desirable, is that its interpretation is not model dependent in contrast to the measures proposed by B&R and S&B. That is, the measure directly employs the predictions on the dependent variable rather than predictions of Y keeping constant the values of the predictors, which changes when the model is adapted.

3. Comparing measures of explained variance: results

In the present section, the measures presented in this paper will be compared in terms of their performances in different scenarios generated explicitly to evaluate their compliance to the two properties defined above: values comprised between 0 and 1, and increased explained variance when adding predictors. A total of six approaches to the estimation of R-squared will be considered:

- the level-specific measures proposed by Bryk and Raudenbush (1992), namely the level-2 modelled variances ($R^2_{2B\&R}$ of equation 14) and level-1 ($R^2_{1B\&R}$ of equation 15), and their sum as measure of total explained variance;

- the level-specific measures proposed by Snijders and Bosker (1992), $R^2_{1S\&B}$ (equation 18) and $R^2_{2S\&B}$ (equation 20), where the former is also interpreted as total explained variance measure;
- the measures of marginal and conditional explained variances proposed by Nakagawa and Schielzeth (2013), $R^2_{N\&S(m)}$ and $R^2_{N\&S(c)}$, respectively equation 22, and 23;
- the measure of overall explained variance proposed by Xu (2003), R^2_{Xu} from equation 25;
- the ANOVA decomposition-based measure proposed in section 2.4, which consists of level specific measures, R^2_{1ANV} and R^2_{2ANV} (equations 26, and 27 respectively), and a measure of total explained variance, R^2_{ANV} (eq. 28);

To test the measures, I generated datasets according to a model-based strategy. The following general equation describes the basic variables involved and their relation:

$$Y_{ij} = \gamma_{00} + \beta_w(X_{ij} - \bar{X}_j) + \beta_b(\bar{X}_j) + u_{0j} + e_{ij} \tag{29}$$

First, a vector of 10 group means, $\bar{X}_j$, is sampled from a defined normal distribution with case-specific parameters. Subsequently, 500 individual-level observations X_{ij} are sampled for each group from a standard normal distribution with group specific means $\bar{X}_j$. The group mean centered transformation of X_{ij} are used to compute the observed outcome values so that the values used for the regression coefficients can be controlled directly by specifying the values of β_w, the within-group regression coefficient, and β_b, the between-group regression coefficient. The error variance term e_{ij} and the group random effects u_{0j} are sampled from normal distributions with mean 0 and case-dependent variances σ^2 and τ^2, respectively, and then added to the computed individual level observations Y_{ij}. Six cases will be discussed in the following. Firstly, in the random intercept context, four cases with datasets generated with the model-based strategy will be discussed. Secondly, a further random intercept case where data was generated with a different approach will be considered. Ultimately, the context of random slopes will be approached with one final example.

3.1 Random Intercept Models

3.1.1 Case 1 – Within- greater than between-group regression coefficient

I set off by first presenting a scenario where we expect all measures to perform similarly and to yield non-negative values between 0 and 1. Data for this first scenario has been generated according to the following equation:

$$Y_{ij} = 0 + 1.3\left(X_{ij} - \bar{X}_j\right) + .7\left(\bar{X}_j\right) + u_{0j} + e_{ij} \tag{30}$$

The 10 group mean values for the variable *X* were chosen to be the 10 integer numbers in the sequence from 0 to 9. Individual observations were sampled from group-specific distribution all with the same variance (set to 1) but with mean corresponding to the group specific mean of *X* just defined. In other words, the observations of the individuals belonging to the same group were sampled from the same normal distribution, with the specific group mean of *X* as mean, and 1 as standard deviation. Once all the observations had been sampled, the mean of the groups were computed to obtain the actual group means. The two regression coefficients β_b and β_w were chosen to be both positive but different so that the averages $\bar{X}_j$ had a smaller effect than the individual level variable X_{ij} had. In particular, $\beta_b =$ .7 and $\beta_w = 1.3$. Finally, the unique effects were sampled from a normal distribution of mean 0 and variance 1, for both the level-1 residuals and level-2 random effects.

The outcome variable was then generated by plugging into equation 30 all the values generated as just described. For a detailed description of the R code used to generate the data see appendix 2.

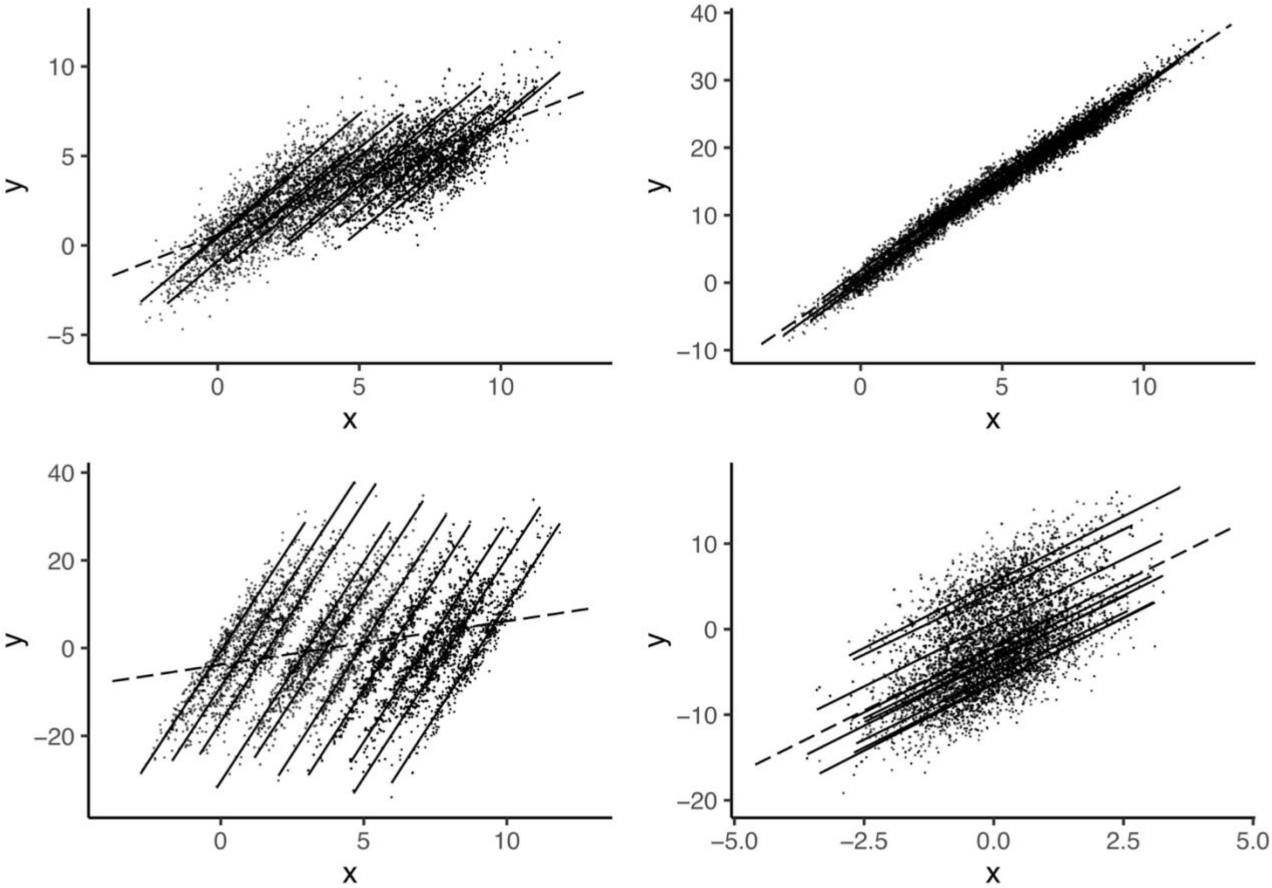

Figure 2 – Scatter plots for the first four datasets are displayed. *Top left*: case 1, within- greater than between-group regression coefficient. *Top right*: case 2, within- equal between-group regression coefficient. *Bottom left*: case 3, between-group regression coefficient close to zero. *Bottom right*: case 4, groups with similar means.

The top left panel in Figure 2 shows a scatter plot of the data and the regression lines estimated by the models. The dashed regression line is the regression line of *Y* on the group averages, while the solid ones are the within group regressions of *Y* on *X*.

Three multilevel models were fitted to the data generated using REML:

A) $Y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$ (the null model);

B) $Y_{ij} = \gamma_{00} + \gamma_{01}\bar{X}_j + \mu_{0j} + e_{ij}$ (a model with the group means as only predictor);

C) $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{1ij} - \bar{X}_j) + \gamma_{01}\bar{X}_j + \mu_{0j} + e_{ij}$ (a model with the within-group deviation score along with the group mean as predictors. Note that this model reflects the true model that was also used to generate all data. Table 3 summarizes the estimated residual variances for the three models for all scenarios. For the current example, the column of interest is the *within greater than between*.

Table 3 – *Variance components estimates for the null model (A), the model with only group means as predictor (B), and the true model (C).*

	within greater than between		within equal between		between approx. 0		similar group means	
Model	$\hat{\sigma}^2$	$\hat{\tau}_0^2$	$\hat{\sigma}^2$	$\hat{\tau}_0^2$	$\hat{\sigma}^2$	$\hat{\tau}_0^2$	$\hat{\sigma}^2$	$\hat{\tau}_0^2$
A. Null model	2.660	3.166	9.749	74.459	104.96	1.787	13.813	17.776
B. Level-2 pred.	2.660	.587	9.749	.573	104.96	1.966	13.813	19.990
C. Level-2 & 1 pred.	1.009	.590	1.009	.590	5.171	2.165	5.038	20.017

As expected, model B, which uses only the group average ($\bar{X}_j$) as predictor, is characterized by the same estimated level-1 unexplained variance as the null model (2.660), and a lower estimated level-2 unexplained variance (0.587). After including the group-mean centered level-1 predictor in the model ($X_{1ij} - \bar{X}_j$) (model C), the estimated error variances in Y is halved. In particular, the estimated residual unexplained variance at the lower level ($\hat{\sigma}^2$) goes from 2.660, in the null model, to 1.009, in the full model; the estimated residual variance at the higher level ($\hat{\tau}_0^2$) decreases from 3.166 to 0.5.

Table 4 shows the computed values for all the R^2 measures considered for model B, the one with only the level-2 predictor. The B&R measure of level-1 explained variance is equal to 0, which means that by adding the group means as explanatory variable the mean squared prediction error is not reduced by any means. The B&R level-2 explained variance is .815 which means that 82% of the group-level variance is explained by the group means (the only predictor in the model). Hence, model B explains *in total* 82% of the variance in Y, according to the B&R measure. All the B&R statistics satisfy the requirements of being greater or equal to zero and smaller or equal to 1.

Table 4 – *R-squared Measures computed for model B estimated on each of the four data examples presented in Figure 2*

	within greater than between			within equal between			between approx. 0			similar group means		
Measure Name	R_1^2	R_2^2	R_{TOT}^2	R_1^2	R_2^2	R_{TOT}^2	R_1^2	R_2^2	R_{TOT}^2	R_1^2	R_2^2	R_{TOT}^2
B&R	0	.815	.815	0	.992	.992	0	-.1	-.1	0	-.125	-
S&B	.443	.815	.443	.877	.992	.877	-.002	-.1	-.002	-.070	-.125	-.070
ANOVA decomp.	0	.997	.517	0	1	.873	0	.822	.014	0	.997	.536
N&S(m)	-	-	.941	-	-	.866	-	-	0	-	-	0
N&S(c)	-	-	.963	-	-	.873	-	-	.019	-	-	.591
Xu	-	-	.423	-	-	0	-	-	0	-	-	0

Moving to the S&B measures, it can be noted that the same holds. The level-2 explained variance is equal to the level-2 B&R measure and can be interpreted in the same way. The level-1 S&B statistic, however, is equal to .443, which suggest that the 44.3% of the variance at the lower level is explained by the group means of *X*. By interpreting $R_{1S\&B}^2$ as a measure of overall explained variance, it can be said that the group means of *X* explain 44% of all the variance in *Y*.

Turning to the ANOVA decomposition measure, it can be noticed that the level-1 explained variance is equal to 0 as was the B&R version, while the measure of level-2 explained variance is practically 1 (.997). However, the measure of total explained variance is more moderate (.557), suggesting that only about half of the overall variance in *Y* is explained by model B, compared to a reference model with no predictors nor random effects. This value is reasonably close to the S&B measure of total explained variance.

The measure proposed by Nakagawa and Schielzeth (2013) suggest that 94% of the overall variance in *Y* is explained by the fixed effects, and that 96% of it is explained by fixed and random effects together. These values are contained in the 0-1 interval but seem to be overly optimistic about the fit of the model.

The measure proposed by Xu (.423) seems to indicate that approximately half of the total variance in *Y* is explained by the group means of X. This value is quite similar to the measures of overall explained variance proposed by S&B and the ANOVA decomposition measure of total explained variance.

Table 5 (left side) shows the summary of the resulting measures of explained variance by model C, using as baseline model A. The level-2 explained variance measure proposed by B&R is equal to the one computed on model B (.814), which means that the inclusion of the within-group variance predictor did not improve the amount of explained variance at the second level. However, the model now explains 62% of the error variance estimated in the null model. These values are in accordance with the first property (0-1 range), while the value of the estimated second level explained variance is actually smaller, by the negligible amount of .001, compared to what was for model B. Furthermore, the measure of total explained variance suitable for the B&R framework well exceeds 1 (1.435).

Table 5 – *R-squared Measures computed for model C estimated on each of the four data examples presented in Figure 2*

	within greater than between			within equal between			between approx. 0			similar group means		
Measure Name	R_1^2	R_2^2	R_{TOT}^2	R_1^2	R_2^2	R_{TOT}^2	R_1^2	R_2^2	R_{TOT}^2	R_1^2	R_2^2	R_{TOT}^2
B&R	.621	.814	1.435	.896	.992	1.889	.951	-.212	.739	.635	-.126	.509
S&B	.726	.814	.726	.981	.992	.981	.931	-.211	.931	.207	-.126	.207
ANOVA decomp.	.621	.999	.817	.897	1	.987	.951	.991	.951	.635	.999	.831
N&S(m)	-	-	.716	-	-	.979	-	-	.931	-	-	.259
N&S(c)	-	-	.821	-	-	.987	-	-	.952	-	-	.851
Xu	-	-	.621	-	-	.879	-	-	.951	-	-	.635

The measures proposed by Snijders and Bosker (1994) follow a similar trend: the level-2 explained variance is practically the same (.815 for model B, .814 for model C) even if it is actually lower by 1 unit at the third decimal place, which is not in line with the property of increasing explained variance; the level-1 explained variance is now .726, which means that by including the group-mean centered level-1 X, the prediction error at the lower level decreases by more than 70%. This can also be interpreted as a reduction of 70% in the value of $\hat{\sigma}_0^2 + \hat{\tau}_{00}^2$, the total variance estimated by the null model.

Moving to the interpretation of the ANOVA decomposition measures, it can be said that 62% of the *within*-group variance is modelled by model C, using a reference to a model without any predictors and without random effects, and 99% of the variance *between* the groups (at the second level) seems to be modelled by model C. The model also explains 82% of the *total* variance in the outcome variable.

According to the measures proposed by Nakagawa and Schielzeth (2013), 72% of the total variance is explained by the fixed effects, and 82% is explained by the fixed and random effects together. The difference between $R^2_{N\&S(m)}$ and $R^2_{N\&S(c)}$ reflects the percentage of the total variability that is due to random effects (here 10%). Finally, according to the Xu measure the residual variation in the response of a person *i* is reduced by 62%, relative to the null model.

Notably, $R^2_{1B\&R}$ yields a measure that is identical to R^2_{1ANV} (.621). The difference with the measure provided by $R^2_{1S\&B}$ is also noticeable. This is due to the different conceptualization of the residual variance provided by Snijders and Bosker; they base the level 1 residual variance on the sum of level 1 and level 2 variance (see equation 16). It is also interesting to notice how the measure of total explained variance proposed in the present article is similar to the measure of total explained variance by both fixed and random effects proposed by Nakagawa and Schielzeth (2013). Finally, the measure proposed by Xu (2003) is identical to R^2_{1ANV}.

3.1.2 Case 2 – Within- equal between-group regression coefficient

The second example dataset was created in such a way that the within and the between-group regression coefficient are the same. The procedure to generate the data followed the one described for the first example, with few changes. In particular, the model used to generate the outcome variable used the same value for both the within-group and the between group regression coefficients ($\beta_w = \beta_b$):

$$Y_{ij} = 0 + 3(X_{ij} - \bar{X}_j) + 3(\bar{X}_j) + u_{0j} + e_{ij} \tag{31}$$

Everything else was left unchanged (see appendix 2 for details). Figure 2 (top right panel) presents a scatter plot of the data generated for the present example.

As in the first example, three models where fitted: null model (A), level-2 predictor only (B), and level-1 and 2 predictors model (C). Table 3 (center left columns, see column headings) presents the estimated unexplained variances by the three models. Results are in line with expectations with the error variance reducing only when the level-1 predictor is added to the model (from model A to model C), and the unexplained level-2 variance reducing just by adding the level-2 predictor (from 74.459 in the null model, to .573 in model B). However, the level-2 residual variance estimated by model C is larger than in model B.

Table 4 reports the estimated R-squared values for model B fitted to this example. The measure proposed by B&R suggests that the model does not explain any variance at the lower level, while it explains 99% of the variance at the higher level. The measure respects the boundaries 0-1 imposed. The level-2 explained variance according to Sniders and Bosker (1994) is also .992. However, the measure of level-1 explained variance suggests that just by adding the group-means as predictors 88% of the variance in individual scores is already modelled, and that the total unexplained variance in Y is reduced by the same amount due to the inclusion of this second level predictor.

The ANOVA decomposition measure indicates that model B is unable to explain any of the within-group variance, all of the between-group variance, and 87% of the overall variance in the outcome variable Y.

The measures proposed by Nakagawa and Schielzeth (2013) indicate that the fixed effects in the model explain 86.6% of the variance in *Y*, and that considering the random effects along the fixed one allows to model 87% of the total variance in *Y*.

The measure proposed by Xu (2003) finds that model B does not explain any variance in the responses of an individual *i*.

It is interesting here to note how the measure of total explained variance proposed in this paper provides exactly the same estimate of explained variance as the one proposed by N&S to account for both the fixed and random effects explanatory power.

Table 5 reports the same statistics just discussed, for the example dataset generated with equal within- and between-group regression coefficients, computed for the linear mixed model C. The inclusion of the within-group deviation predictor (level-1) in the model, along with the group-means of X, models 89.6% of the variance at the individual level, according to the B&R leve-1 explained variance measure. The level-2 explained variance measure is unchanged, .992. This means that while the level-specific measures are not in contrast with the 0-1 property, the measure of total explained variance by the model greatly exceeds 1 (1.889).

According to the measures computed under the guidelines of S&B, 98% of the variance at the individual level, and of the overall variance in *Y*, is explained by the addition of the group-mean-centered level-1 predictor *X* and the group means of *X*. The amount of variance in *Y* explained at the second level is the same that was already explained by model B.

The ANOVA decomposition measure follows the same tendency: model C explains almost all the within-group variance (90%), it explains all the variance at the second level, and, compared to the model with only the group-means as predictors, it explains more of the overall variance in *Y* (99%).

Following the N&S measures we can say that in model C, the random effects do not explain much of the total variance in the outcome variables, as the difference between the *marginal* (.979) and *conditional* (.987) measures is rather contained. To conclude, the estimation proposed by Xu

(2003) indicates that the inclusion of both the level-1 and 2 predictors explains 88% of the individual variance in *Y*.

All measures of explained variance considered for model C are in line with the expected range for an R-squared statistics of 0-1, except for the measure of total explained variance proposed for the B&S computational framework. For all measures it is also true that adding a predictor leads to a greater (or equal) estimated modelled variance.

As for model B, in the case of the data generated by a model with the same within- and between-group regression coefficient, the measures R^2_{ANV} and $R^2_{N\&S(c)}$ are identical.

3.1.3 Case 3 – Between-group regression coefficient close to zero

The data used in the next example is similar in structure to the example described in section 2.2.3. As in that case, the between-group regression coefficient is close to 0. In particular, the model used to generate the dataset is characterized by a between-group regression coefficient equal to 0 ($\beta_b = 0$):

$$Y_{ij} = 0 + 10\left(X_{ij} - \bar{X}_j\right) + u_{0j} + e_{ij} \tag{32}$$

The data generation procedure follows the same structure that was presented for cases 1 and 2 (the code can be reviewed in appendix 2). However, apart from the between-group regression coefficient, the within-group regression coefficient is larger, set to 10, as it is the error variance and variance of the random effects, now set to 5. Figure 2 (bottom left panel) presents the scatterplot of this data.

By fitting the usual null model (A), it is expected that the second level variance estimate will be relatively low. By adding the group mean as a predictor to the model, it is expected that both the level 1 and 2 residual variances will remain unchanged (model B). Finally, the addition of both the group means variable and the within-group deviation variable is expected to reduce significantly both variances (model C).

Table 3 (see column headers) reports the residual variances. While $\hat{\sigma}^2$ remained unchanged when $\bar{X}_j$ was added (104.96) and decreased when X_{ij} was added (to 5.171), as one would expect, $\hat{\tau}_0^2$ increased as explanatory variables were added to the null model. Adding the level-2 predictor led $\hat{\tau}_0^2$ to increase from 1.787, to 1.966, and adding both the level-1 and level-2 predictors (model C) led to an even higher increase of the estimated unexplained second level variance (2.165). As a consequence, neither R&B nor S&B can provide a meaningful measure of modelled variance at the second level. In fact, for both measures to provide values that are comprised between 0 and 1, it is necessary that the estimated unexplained variance does not increase as predictors are added.

Table 4 also shows the R-squared measures computed for model B fitted to the dataset generated with between-group regression coefficient approximately 0. The formulas proposed by B&R provide a measure of modelled variance that is practically 0 for both the level 1 and level 2 variances. As expected, it the level 2 explained variance measure is actually negative. The measures computed according to S&B also provide estimated explained variances that are practically zero but numerically negative.

The ANOVA decomposition measure is in line with the previous two when calculating the explained variance within the groups and in the overall variance of Y (0 and .014 respectively). However, the measures are not negative, and therefore more in line with the desired properties. Furthermore, the measure seems to indicate that the inclusion of the group-means of X as a predictor in the model, and the specification of the random effects, leads to a proportionate reduction in the prediction error at the second level of more than 80%.

According to all the N&S, and Xu measures, the model does not explain any variance in Y.

Table 5 shows the computed values of all of the R-squared measures for model C when applied to this third example. For what concerns the level-1 explained variance, 95% of the variance at the lower level is explained by the addition of the group means and the group-mean centered X to the null model, according to the B&R measure. The S&B level-1 measure provide a similar measure with an estimated 93% proportional reduction in the estimated mean squared prediction error, compared to the null model. However, both R&B and S&B level-2 explained variances are negative as anticipated. As a consequence, they are not interpretable.

The ANOVA based measure provides interpretable results describing a high explanatory power of the model at both levels ($R^2_{1ANV} = .753$, $R^2_{2ANV} = 1$), and overall ($R^2_{ANV} = .959$). However, the value of 1 for the second-level explained variance measure with the R^2_{2ANV} was not anticipated and it demands further investigation. All the other total explained variance measures are quite similar, and Xu is again equal to the level-1 B&R and ANOVA level-1 measures.

Finally, by comparing Table 4 with Table 5, the behavior of the measure when the within-group deviation score predictor is added can be assessed. In particular, it is interesting that for the dataset generated with $\beta_b = 0$ and $\beta_w = 10$, $R^2_{2S\&B}$ and $R^2_{2B\&R}$ are not only negative for both model B and C, but they also decrease from model B to C (where the latter has more predictors than the former). On the other hand, the level-2 ANOVA measure of modelled variance sees an increase in its value when the level-1 predictor is added to the model, and in this sense it provides a measure of explained variance more in line with the desired features of an R-squared measure.

3.1.4 Case 4 – Groups with similar means

For this fourth scenario, I randomly sampled the group means from a normal distribution with mean 0 and standard deviation .01. In other words, all groups have rather similar X_{ij} mean values but differ in their average Y value (see Figure 2 bottom right panel for the scatterplot). The variance used to sample the random effects was also changed to 10. This was done to create more variance among the group intercepts. Had this not been done, the groups would have been extremely similar. The detailed code used to generate the data is presented in appendix 2.

The three models used in the previous example were fitted again to the data generated for this example: a null model (A), a model with only the group means as predictors (B), and a model with both the group means and the within-group deviation score as predictors (C). Table 3 reports the estimated residual variances for the three models in the block of columns to the right. The level-1 residual variance decreases only from model A to model C (due to the inclusion of the level-1 predictor) while the level-2 variance increases both from model A to B and from model A to C.

Table 4 reports in the last block of columns all the R-squared measures of interest computed on model B. According to the B&R measure of level-1 explained variance, the model with only the group-means as predictor does not explain any variance in the individual responses of *Y*. The second level explained variance is negative and therefore not interpretable. As a result, the model should explain a negative total amount of variance, which does not have any meaning. Similarly, all measures computed according to the S&B formulas are negative and uninterpretable.

The explained variance computed according to the ANOVA decomposition measure is 0 for what concerns the first level, which means that the model with only level-2 predictors and random effects does not explain any variation within the groups. However, according to this computational framework, the model explains all the variation between the groups (99%) and a little more the half of the variance in the total variance of Y (54%).

Looking at the N&S measures it seems that all the variance that the model explains is attributable to the random effects, since the *marginal* version of this measure is 0 and the *conditional* version is .591. The measure proposed by Xu indicates a complete lack of fit of the model.
Interesting to notice is that once again the measure measures of total explained variance proposed in this paper is fairly close to the conditional R-squared proposed by N&C.

Table 5 shows the computations for the same measures as applied to model C, with the null model as reference. Adding the group-mean centered version of the level-1 predictor and the group-means as predictors, helps to explain 64% of the variance in the individual level responses, according to B&R measure. However, at the second level an uninterpretable value of explained variance is found once again for the measure of level-2 explained variance proposed by B&R. The same holds for the measure level-2 measure proposed by S&B. According to the S&B approach, adding the two predictors that characterize model C leads to a proportionate reduction in the total variance in *Y* of .207.

The ANOVA measure points toward a much greater explained variance by the mode both within the groups (.635) and overall (.831). According to this computational framework, the model explains all the between-group variance (99%).

Moving to the measures proposed by Nakagawa and Schielzeth (2013), it seems that over 60% of the explanatory power of the model is due to the random effects.

Finally, 64% of the individual variance in the responses of each individual is explained by the model with both predictors according to the Xu measure.

The similarities highlighted in previous examples are found again: R^2_{1ANV} is equal to $R^2_{1B\&R}$ (and R^2_{Xu} and $R^2_{V\&C}$, .635), and the measure of total modelled variance R^2_{ANV} (.831) is approximately equal to the one proposed by Nakagawa and Schielzeth (2013), which is .851.

3.1.5 Case 5 – Low explanatory power for the predictor group means

A last example dataset for the random intercept setting is considered. A scenario where the group means of the response variable cannot be explained, neither by the within nor with the between group relations. The idea is to have groups with the same group X_{ij} means, but different Y group means, so that the predictor group means have low explanatory capability.

To generate this dataset, a different approach was used. In each of 10 groups of 500 observations, the individual scores of X and Y were sampled from a normal multivariate distribution with variance-covariance matrix $\sum = \begin{bmatrix} 10 & 7 \\ 7 & 10 \end{bmatrix}$. Five values were selected for the means used for the sampling of the individual values of X: 0, 10, 20, 30, and 40. From the 10 groups, five couple were defined. In each couple of groups, the same group mean was used to sample the values of X.

As for the values of Y, a similar approach was followed. The 10 groups were divided into two sets. One set of five groups was assigned a Y mean of 5 and the other set was assigned a Y mean of 15. For a detailed description of the R code used to generate the data, see appendix 2. Figure 3 shows the data pattern.

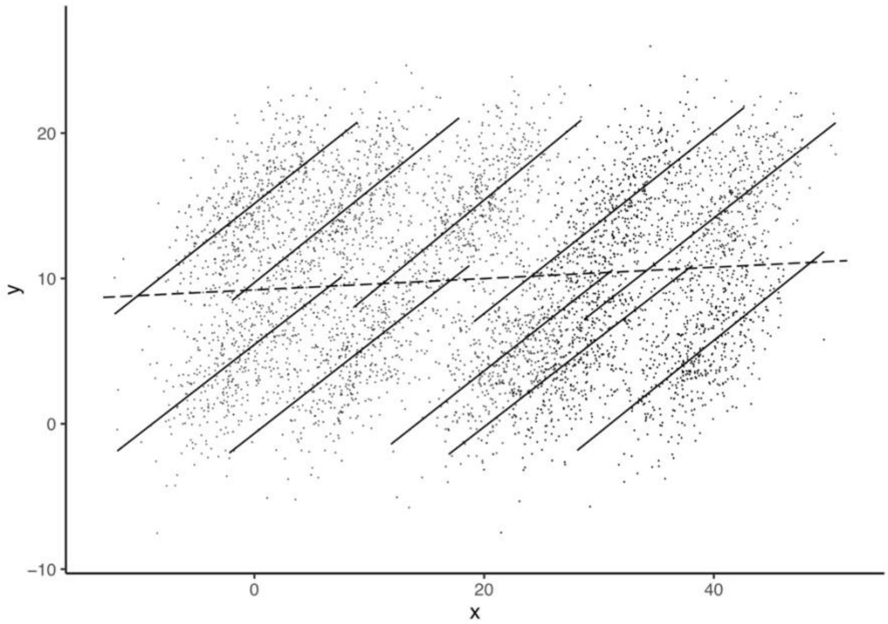

Figure 3 – *Scatterplot of the dataset generated based on a normal multivariate distribution.*

As in the previous cases, three models are of interest: the null model, a mixed model with only the group means as predictor (model B); and a mixed model with both group-means and individual values of *X* as predictors (model C). Table 6 presents the estimated residual variances by these three models. Adding the group-level variable leads to 0 the estimated unexplained level-2 variance and has a minor effect on the estimated error variance (10.634, instead of 10.651). The behavior of the estimated unexplained variance is in line with what one would expect when adding both the level-2 and level-1 predictor are added. However, the estimated unexplained level-2 variance in model C is strikingly higher than in model A (125.739 compared to 27.693, respectively).

Table 6 – *Variance components estimates for the null model (A), the model with only group means as predictor (B), and the true model (C).*

Model	$\hat{\sigma}^2$	$\hat{\tau}_0^2$
A. $Y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$	10.651	27.693
B. $Y_{ij} = \gamma_{00} + \gamma_{01}\bar{X}_j + \mu_{0j} + e_{ij}$	10.634	0
C. $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{1ij} - \bar{X}_j) + \gamma_{01}\bar{X}_j + \mu_{0j} + e_{ij}$	5.705	125.739

Table 7 (left side) reports the estimated R-squared measures for model B. According to the measures proposed by B&R the model does not explain any variance at the lower level and it explains all the variance at the higher level. However, the measure of overall explained variance exceeds 1, even if by a negligible amount (1.002).

According to S&B level-1 explained variance, adding the group-means as predictor, leads to a substantial (72%) proportional reduction in the value of $\hat{\sigma}^2 + \hat{\tau}_0^2$, the total variance in Y, which can also be interpreted as the contribution of the predictor to the explained variance at level-1. The level-2 explained variance is equal to unity, suggesting that the model explains all the variance between the groups.

The ANOVA decomposition measure suggests that model B is unable to model any of the within-variance but all the between-group variance. The total amount of variance explained is estimated to be 70%. The same value is estimated with both the measures computed with the N&S strategy.

Finally, the Xu measure indicates that the group-mean variable is unable to model any of the variation in the response of a person *i*.

When it comes to model C, Table 7 (right side), the B&R measure of level-1 explained variance indicates that the addition of both the group-means and the within-group deviation score (X_{ij}) models 46% of the variance at the individual level. The measure of level-2 explained variance is not interpretable, being it a negative value. The same problem is found with $R^2_{2S\&B}$ (-3.54). The measure proposed by S&B is uninterpretable also at the level-1 (-2.428).

The ANOVA decomposition measure does, however, estimate that the model explains 46% of the within-group variance, and 84% of the total variance in *Y*. The estimate of the level-2 variance provided by this type of measure is again 1. Nevertheless, this measure is the first to provide values between 0 and 1 for all the considered variances.

Considering the measures proposed by N&S it seems that the fixed effects explain almost half of the variance in Y ($R^2_{N\&S(m)} = .489$) and that the random effects explain almost all the residual variance left ($R^2_{N\&S(c)} = .978$).

Finally, the measure proposed by Xu (2003) indicates that model C explains 46% of the variation in the responses of an individual.

All of the ANOVA measure, the Xu and the conditional R-squared measures increment the estimated value of explained variance when the level-1 predictor is added, while the other see a decrease.

Table 7 - *R-squared measures for model B and C fitted to the dataset generated for the fifth case (null model A as reference)*

	Model B			Model C		
Measure Name	R^2_1	R^2_2	R^2_{TOT}	R^2_1	R^2_2	R^2_{TOT}
B&R (1992)	.002	1	1.002	.464	-3.54	-3.076
S&B (1994)	.723	1	.723	-2.428	-3.54	-2.428
ANOVA decomposition	0	1	.701	.463	1	.839
N&S (m)	-	-	.701	-	-	.498
N&S (c)	-	-	.701	-	-	.978
Xu (2003)	-	-	0	-	-	.464

3.2 Random Intercept and Random Slope Models

To test the measurement performances on models specified with random slopes, a dataset is generated with a model-based strategy as in cases 1 through 4. The main difference with the previous case is that random slopes are used along with random intercepts in model 29. In particular, β_w is not fixed anymore, but the result of a fixed (shared) coefficient, γ_{00} (here chosen to be 5), and a random part, u_{1j}. For this example, the random group effects u_{0j} and u_{1j} are sampled from a multivariate normal distribution with covariance matrix $\sum = \begin{bmatrix} 5 & 4.5 \\ 4.5 & 5 \end{bmatrix}$, where both the variance of the random intercept τ_0^2 and of the random slopes τ_1^2 is 5 (diagonal elements), and their covariance τ_{01} is 4.5 (off-diagonals). The fixed between-group effect used to create the dataset is $\beta_b = 2$.

As for the level-1 explanatory variable, its values are sampled from a normal distribution with variance 1 and mean corresponding to the group id the observations belong to (individual values of X_{ij} belonging to different groups were sampled from different distributions with group specific means). Finally, Y is generated according to the model in equation 29. The full account of the generation procedure is accompanied by the R code in appendix 2. Figure 4 presents the scatterplot of the data just described.

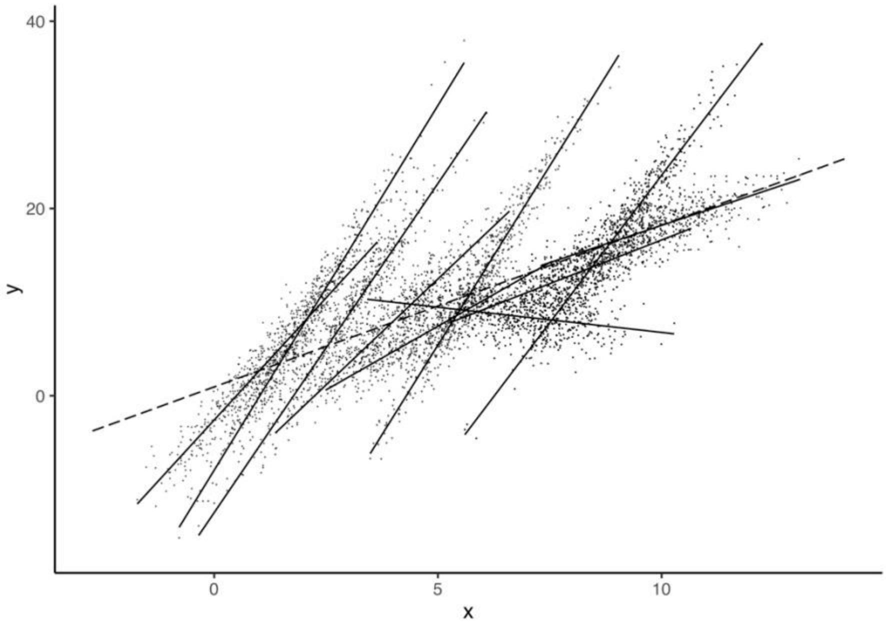

Figure 4 – *Scatterplot of the dataset generated with random intercepts and slopes.*

Along with the null model (A), two more models were fitted to the dataset: one with both level-1 and 2 predictors but specifying only the intercepts as random (B); one with both predictors and both random intercepts and random slopes (C).

The first model (B) was specified because the variances estimated by it are needed to compute the B&R and S&B measures, which, according to Snijders and Bosker (1994, 1999), should be computed on a random intercept model even when the researcher has sensible reasons to believe that the slopes are not constant among groups. The reason is that, as they claim, the more complicated computations of their measures when adapted to a random slope case provide the same estimated explained variances that are computed on an equally specified model that ignores the random slopes. Table 8 presents the estimated residual variance by model B on the data generated for this specific case. As can be seen the addition to the model of both the within-group deviation score and the group-means, reduced the estimated unexplained variances at both levels.

Table 8 – *Random Intercept Model for Residual Variances in* $R^2_{*S\&B}$ *and* $R^2_{*B\&R}$

Model	$\hat{\sigma}^2$	$\hat{\tau}_0^2$
A. $Y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$	29.056	23.484
B. $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{1ij} - \bar{X}_j) + \gamma_{01}\bar{X}_j + \mu_{0j} + e_{ij}$	10.461	5.643

Model C, on which the other explained variances measures were computed, is:

$$Y_{ij} = \gamma_{00} + (\gamma_{10} + u_{1j})(X_{ij} - \bar{X}_j) + \gamma_{01}(\bar{X}_j) + u_{0j} + e_{ij}$$

Table 9 reports the estimated R-squared measures. The S&B measure of level-1 explained variance, computed on model B but interpreted for model C, affirms that the predictors added to the null model explain 70% of the variance at level-1, and therefore that the estimated residual total variance is 70% smaller when estimated by the mixed effect model. 76% of the variance at the second level is modelled by model C according to both the S&B and B&R measures of level-2 explained variance. At level 1, the B&R measure estimates that 64% of the variance in Y is explained by the model. However, the total amount of explained variance estimated according to the B&R framework well exceeds 1.

Moving to the explained variance estimated with the ANOVA decomposition measure, it seems that adding both level-1 and level-2 predictors and the random effects models 90% of the within-in group variance, and 94% of the total variance in Y. As for the variance at the second level, it is estimated that the model explains all of it.

In this case, the measures of explained variance proposed by Nakagawa and Schielzeth (2013) were not computed as there is not a straight forward way to apply them to a random slope scenario. Finally, Xu's measure indicates that 90% of the variance in individual reposes is explained by model C.

Table 9 – *Random Slope Model R-squared measures*

Measure Name	R_1^2	R_2^2	R_{TOT}^2
Bryk and Raudenbush (1992)	.640	.760	1.400
Snijders and Bosker (1994)	.693	.760	.693
ANOVA decomposition	.896	1	.940
Xu (2003)	-	-	.896

4. Discussion

The present article addressed the concept of explained variance as a measure of goodness of fit in the context of linear mixed models. The topic was approached by reviewing measures established in the literature and comparing their performances in six fictitious scenarios. An ANOVA decomposition-based way to compute measures of both overall and level-specific explained variances has been proposed and tested along with the other measures reviewed. Overall, five measuring approaches were compared: the level-specific measures of modelled variance proposed by Bryk and Raudenbush (1992); the level-specific ones proposed by Snijders and Bosker (1994); the measures of total explained variance proposed by Nakagawa and Schielzeth (2013); the measure of overall modelled variance proposed by Xu (2003); and the ANOVA decomposition-based measure of within-group, between-group, and total modelled variances. The main focus in evaluating the performances of the measures was to consider whether they abided by two properties recognized in the literature as desirable for an R-square measure (Kvalseth, 1985): the estimated value must be included in the range 0-1, and it must increase as predictors are added.

The six cases presented in the paper contribute in building a body of evidence that shows how the most established measures in the literature (Bryk and Raudenbush (1992), and Snijders and Bosker (1994)) do not generally comply with the two desirable properties of an R-square measure. Contrarily, in all the cases examined, it was found that the proposed ANOVA decomposition measure never fell outside the interval of meaningful R-squared values and never decreased as predictors were added. In this regard, the approach proposed here is a compelling way to estimate the explained variance by linear mixed models.

The issue of negative R-squared values that characterizes the computation strategy proposed by Bryk and Raudenbush (1992) is well known in the literature and Snijders and Bosker (1994, 1999) already addressed it diffusely. The level-2 explained variance measure can be negative because in the approach proposed by B&R the unexplained level-2 variance is estimated exclusively by estimating the variance of the random intercepts, τ^2. This feature leads to negative R-squared estimation in the not so infrequent case in which adding predictors to a mixed model will increase the estimated $\hat{\tau}^2$. Furthermore, the measure of total explained variance defined as the sum of the B&R explained variances at level-1 and 2, easily yields values greater than one. Both these behaviors were found in the examples presented in this article (see case 1 and 2 for the overall negative values, and case 3 and 4 for the level-2 negative values).

The S&B measure resolves the issue of negative overall explained variance, and in fact it was not found to be negative for any of the cases considered. However, in case 3 and 4, when data were generated with no between group effect and with similar group means of X, respectively, the measure of explained variance at level-2 did yield negative values. Snijders and Bosker (1994) already discussed these possibly negative values and actually claimed it to be a desirable feature for their R-squared measure at level-2. According to their account, their $R^2_{2S\&B}$ formula (equation 20) may estimate negative explained variances only when the within- and between-group effects are wrongly forced to be equal by a particular model specification. Hence, they claim that their measure has the positive feature of diagnosing model misspecifications. We dispute their claim, as in our third and fourth case we did not force the within- and between-group regressions to be equal and still found negative level-2 explained variances measures. The same undesirable behavior was found in case 5, where once again the model did allow for different within- and between-group effects. Hence, clearly, disproving the misspefication of the model to be the only cause of estimation of negative explained variances at the second level, with the S&B formula. Another issue, with the claim for model-diagnostic capabilities of the S&B level-2 measure, is that all models except the *true* model are misspecified. Hence, evidences of misspecification may not be as useful as Snijders and Bosker (1994) uphold. A final remark on the performance of the S&B level-2 explained variance measure is that, in cases 3 and 4, it was found to be not only negative, but also to yield a smaller value when adding the level-1 group-mean centered predictor (going from model B to C). This is not ideal since it violates the second property of non-decreasing explained variance. To conclude, a measure that behaves according to the properties defined may be more valuable than a measure that provides evidence of a phenomenon which is already known.

The ANOVA decomposition approach to the estimation of explained variance for multilevel models is a comprehensive method to estimate modelled variances both in level-specific and overall terms. It proved to yield values in the defined interval 0-1, and increased explained variance estimations when variables are added to models. Another positive feature of the ANOVA decomposition measure here proposed is that it applies directly as it is to random intercept models as to random slope models. In this regard it has a clear edge on the other measures which provide information on the model fit either indirectly (R&B and S&B) or after some adjustments (Jonson (2014) version of Nakagawa and Schielzeth (2013) measures). Its level-1 measure is interpretable as the percentage of within-group variance that is modelled by the predictors, and the overall measure is interpretable as the amount of the total variance in *Y* that is accounted for by the model specification. The measure proposed to estimate the level-2 explained variance yielded a value close to unity in all examples. The reason for this is that it includes the level-2 random effects, i.e., the random effects are counted as explained variances. The measures proposed by Bryk and Raudenbush

(1992) and by Snijders and Bosker (1994) do not. Since we believe that the random effects at level-2 should be excluded from the computation, we conclude that the level-2 explained variance of the ANOVA approach is not meaningful.

Some similarities have been noticed among the measures in the values that they yield. In particular, it seems that the measure proposed by B&R to estimate the modelled variance at the first level is equal to the one computed according to the ANOVA framework in all the random-intercept cases considered (cases 1-5). In the same cases, level-1 ANOVA decomposition-based measure is also equal to the one proposed by Xu (2003). These similarities strengthen the use of the ANOVA decomposition measure as a measure of level-1 explained variance.

The general agreement between the measure of total explained variance proposed in the ANOVA framework and the measure of conditional R-squared proposed by Nakagawa and Schielzeth (2013) is also interesting. Both measures consider the contribution of the random effects to the modelling of variance and in this might lie the reason of such similarities. However, further research might explore the link between the two.

The conclusions and discussion points just put forward were drawn from the comparison of measures performances tested on only six fictitious cases. These are not comprehensive of all the possible scenarios in which these measures are applicable, and the random slopes context has been oversighted in favor of the random intercept one. Therefore, general conclusions on the advantages offered by the ANOVA decomposition measure on the more stablished measures should be made with caution. Ideally, one may prove properties of all measures (for instance, under which conditions they satisfy the two desirable properties of an R-square measure), and provide results on their properties in more general conditions than examined in the present paper. Moreover, some problems of the measures might be theoretically interesting to investigate but might not be particularly relevant in practice. In particular, the conditions in which the S&B and B&R estimates of level-2 explained variance result into negative values seem to be somewhat related to the fact that the unexplained variance at level-2 is high, compared to one at level-1, in most of the data generated. In practice, it is not clear how often these situations may occur.

Future research might try to compare the measures on different datasets (including real ones) and to adjust the measure of level two explained variance in the ANOVA computational framework to make it more meaningful. The modification of the ANOVA measure could be pursued by changing the predicted value of the group means ($\bar{y}'_j$) in the numerator of equation 27. Since the problem lies in the fact that the measure incorporates the level-2 random variance in the explained variance, one could for example try to substitute it with predictions done only based on the fixed effects, but this and other options may be examined further. Another direction for improving the level-2 explained variance of the ANOVA framework would be to act on the implicitly defined null model in the denominator. For example, instead of the observed group means minus the observed grand mean, one could think of using the predicted group means by a random intercept only model in the denominator of equation 27. However, to be consistent with the other ANOVA measures, this would require changing this also in the measures of level-1 and total explained variance, which we do not recommend.

References

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences.*

Cameron, A. C., & Windmeijer, F. A. (1996). R-Squared Measures for Count Data Regression Models With Applications to Health-Care Utilization. *Journal of Business & Economic Statistics, 14*(2), 209-220.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis.* New York: NY:Wiley.

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944-946.

Kvålseth, T. O. (1985). Cautionary Note About R2. *The American Statistician, 39*(4), 279-285.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Sage.

Snijders, T., & Bosker, R. (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research, 22*(3), 342-363.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and applied multilevel analysis.* London: Sage.

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. *International Biometric Society, 52*(2), 572-587.

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in m edicine, 22*(22), 3527-3541.

Appendix 1 – Commented R code for R^2 measures computation

Here I present all the measures implementations in R.

Bryk and Raudenbush (1992)

The first step to compute the B&R measure is to extract from the null and full models the estimated residual variances.

```
Null.Model <- lmer(y ~ 1 + (1|grpid), data = any.data)
Model.1 <- lmer(y ~ (Xij_gmc) + Xj + (1|grpid), data = any.data)# (e.g
. model C)

sigma2 <- as.data.frame(summary(Model.1)$varcor)[2, "vcov"] #lvl-1 une
xplained vairance (full model)
sigma2_n <- as.data.frame(summary(Null.Model)$varcor)[2, "vcov"] #lvl-
1 unexplained vairance (null model)

tau2 <- as.data.frame(summary(Model.1)$varcor)[1, "vcov"] #lvl-2 unexp
lained vairance (full model)
tau2_n <- as.data.frame(summary(Null.Model)$varcor)[1, "vcov"] #lvl-2
unexplained vairance (null model)
```

Subsequently, the computations described in equation 14 and 15

```
R2_1_BR <- 1-(sigma2/sigma2_n)  # lvl-1 explained variance
R2_2_BR <- 1-(tau2/tau2_n)      # lvl-2 explained variance
R2_T_BR <- R2_1_BR + R2_2_BR    # total explained variance
```

Snijders and Bosker (1994)

The computation of Snijders and Bosker measures requires the same variance estimation computed just above, plus the computation of the harmonic mean of group sizes.

```
# Harmonic Mean
N.tot <- nrow(any.data) #total number of cases in the dataset
den.hm <- rep(NA, length(group_v)) # denominator for the harmonic mean
  for(d in 1:length(den.hm)){
    if(nrow(any.data[grpid == group_v[d], ]) > 0){
      den.hm[d] <- 1/nrow(any.data[grpid == group_v[1], ])
    }
    else {den.hm[d] <- 0}
  }
hm <- N.tot/(sum(den.hm)) # computation of the harmonic mean
```

I then computed the measures as follow:

```
R2_1_SB <- 1-((sigma2 + tau2)/(sigma2_n + tau2_n))        # lvl-1 expl
ained variance
R2_2_SB <- 1-((sigma2/hm + tau2)/(sigma2_n/hm + tau2_n))  # lvl-2 expl
ained variance
R2_T_SB <- R2_1_SB                                        # total expl
ained variance
```

Nakagawa and Schielzeth (2013)

```
# > Marginal ####
  # Sf: variance attributable to fixed effects
    X <- model.matrix(Model.1)
    Beta <- fixef(Model.1)
    Sf <- var(X %*% Beta) # the variance of the predicted values based
 exclusively on the fixed effects
  # Sl: netsing unit effects variance
    Sl <- tau2
  # Se: error variance
    Se <- sigma2
  # R2 computation
    R2_NSm <- Sf / (Sf+Sl+Se)
# > Conditional ####
    R2_NSc <- (Sf+Sl) / (Sf+Sl+Se)
```

Where `Model.1` is the full mixed model, and `sigma2` and `tau2` are extracted from it as in the computation of the B&R and S&B measures.

Xu (2003)

To compute the measure proposed by Xu (2003), the predicted values estimated by the null and full model are extracted and then used to compute the residual sum of squared errors. The important detail here is that the package `mbset` was used to estimate the model. The function `mhglm_ml` in this package predicts the random effects using the empirical Bayes instead of the conditional modes, as with the function `lmer` from packages such as `lme4`.

```
# Model estiamtion with mbest pacakge
Null.Model_EB <- mhglm_ml(y ~ (1|grpid), data = any.data)
Model_EB <- mhglm_ml(y ~ (Xij_gmc) + Xj + (1|grpid), data = any.data)

#Full model RSS
  nij.p <- fitted(Model_EB) #predictions under the mixed model
  RSS <- sum((y-nij.p)^2)
#Null model RSS
  nij0.p <- fitted(Null.Model_EB) #predictions under the null model
  RSS0 <- sum((y - nij0.p)^2)
```

```
#R^2
  R2_Xu <- 1 - RSS/RSS0
```

ANOVA decomposition measure

First, the level-1 explained variance:

```
#Group means of y (need for both numerator and denominator)
  yj <- NA
  yj_unique <- aggregate(y, list(grpid), mean)
  for(i in 1:nrow(any.data)){
    yj[i] <- yj_unique[yj_unique$Group.1 == grpid[i], 2] #vector of ob
served outcome group means (each case has its group mean as value)
  }
#SSw_P (Numerator)
  SSw_P <- sum((yj-predicted)^2)
#SSw (denominator)
  SSw <- sum((y-yj)^2)
#R1^2
  R1_2 <- SSw_P/SSw
```

Then, level-2

```
# Preidcted group means
  predicted_gm <- aggregate(predicted, list(grpid), mean)[, 2]
  yj <- aggregate(y, list(grpid), mean)[, 2]
#B (numerator)
  SSb_P <- sum((predicted_gm - mean(y))^2)
#SSb (denominator)
  SSb <- sum((yj - mean(y))^2)
#LVL2 R-squared measure
  R2_2 <- SSb_P/SSb
```

And finally, the measure of total explained variance

```
#TSS with predicted
  TSSp <- sum((predicted-mean(y))^2)
#TSS
  TSS <- sum((y-mean(y))^2)
#TOT R-squared
  Rtot <- TSSp/TSS
```

Appendix 2 – Commented R code for data generation

Code and detailed description for the data generation process for each case follows.

Case 1 – *Within- greater than between-group regression coefficient*

First of all, all the parameters are defined:

```
# GROUP DETAILS
  Ng <-          10  #number of groups
  Nj <-          500 #number of members
# VARIANCEs
  Xj_var <-      1 #muliplier term to make Xj variance big or small
  Xij_var <-     1 #sampling sd for Xij generation
  lvl1unexp <-   1
  lvl2unexp <-   1
# GAMMAs
  gamma00 <-     0   #Expected Y-grand mean (intercept for the average 
group)
  gamma10w <-    1.3 #Within group effect
  gamma01b <-    .7  #Between group effect
```

Subsequently, these parameters are used in the following code to actually generate the dataset. The first step is to create a vector of group-id values for each of the observations. This is done based on the number of groups and observations within groups defined above.

```
grpid <- rep(seq(1, Ng), each = Nj)
```

Next, the values of the predictor X are computed. First, a vector of group means is generated based on the group identification code. `Xj_var` is set to 1 in this example, so it can be ignored.

```
#LVL-1 predictor: any continuos random variable (Xij)
  #Note: Xij mean depends on group membership
  Xj_gen <- seq(0, max(Xj_var*unique(grpid)), by = Xj_var) # vector of
 group means from which to pick the Xij
```

Second, from each group mean, 500 observations are drawn based on the standard deviation `Xij_var` defined above.

```
set.seed(180524)
Xij <- vector("list", Ng)
  for(g in 1:max(grpid)){
    Xij[[g]] <- rnorm(Nj, Xj_gen[g], Xij_var)
    # changin Xij_var changes the total level 1 variance (null model e
stimate)
    # but keeps the residual varaince estiamted in model 1 the same
  }
  Xij <- stack(as.data.frame(do.call("cbind", Xij)))[, 1] # puts all o
bservations together in one vector
```

At this point, a vector of actual group mean values of X is generated. A mean is computed among the members of each group and subsequently attributed to each observation. As a result, an `Xj` vector exists for which all individuals in the same group have the same value. The group mean centered version of `Xij` (called `Xij_gmc`) is created right after.

```
#LVL-2 predictor: Xij group means (Xj)
  Xj <- rep(aggregate(Xij, list(grpid), mean)[, 2], each = Nj)
#Group mean centered Xij
  Xij_gmc <- Xij-Xj
```

Next, the unique and random effects are generated by sampling from a normal distribution with mean 0 and standard deviation the square root of `lvl1unexp` and `lvl2unexp`, which were defined at the beginning and represent the target level-1 and level-2 unexplained variances (a value very close to these parameters will come up when estimating the residual variances with the random intercept model with predictors).

```
#Random effects
  eij <- rnorm(Ng*Nj, 0, sqrt(lvl1unexp)) #individual level variance a
round the group reg line
  u0j <- rep(rnorm(Ng, 0, sqrt(lvl2unexp)), each = Nj) #random group m
ean variance around the grand mean
```

Finally, the outcome variable is generated. To do so we first define the regression coefficients, then y is generated according to the defined generating model (equation 30), and ultimately all the variables of interest are put together in a data frame.

```
#Level 2 equations
  B0 <- gamma00 + u0j
  Bw <- gamma10w
  Bb <- gamma01b
#Level 1 equation (Note: with values inserted it becames the combined
model)
  y = B0 + Bw*(Xij-Xj) + Bb*(Xj) + eij
  yj <- rep(aggregate(y, list(grpid), mean)[, 2], each = Nj) # group m
eans of Y
  any.data <- as.data.frame(cbind(grpid, y, Xij, Xj, Xij_gmc = (Xij-Xj
), eij, u0j))
```

Case 2 – *Within- equal between-group regression coefficient*

Here the code is presented as a whole. For the break-down comments see the code used in case 1. The only differences compared to the previous case are the values of the `gamma10w` and `gamma10b` (the within- and between-group regression coefficients) which are now both equal to 3.

```
# GROUP DETAILS
  Ng <-          10 #number of groups
  Nj <-          500 #number of members (try with larger samples)
# VARIANCEs
  Xj_var <-      1 #muliplier term to make Xj variance big or small
  Xij_var <-     1 #sampling variance for Xij generation
  lvl1unexp <-  1
  lvl2unexp <-  1
# GAMMAs
  gamma00 <-     0 #Expected Y-grand mean (intercept for the average gr
oup)
  gamma10w <-    3 #Within group effect
  gamma01b <-    3 #Between group effect


# DATA GEN
  grpid <- rep(seq(1, Ng), each = Nj) # group ids
  #LVL-1 predictor: any continuos random variable (Xij)
    #Note: Xij mean depends on group membership
    Xj_gen <- seq(0, max(Xj_var*unique(grpid)), by = Xj_var) # vector
of group means from which to pick the Xij
    set.seed(180524)
    Xij <- vector("list", Ng)
      for(g in 1:max(grpid)){
        Xij[[g]] <- rnorm(Nj, Xj_gen[g], Xij_var)
      }
      Xij <- stack(as.data.frame(do.call("cbind", Xij)))[, 1]
  #LVL-2 predictor: Xij group means (Xj)
    Xj <- rep(aggregate(Xij, list(grpid), mean)[, 2], each = Nj)
  #Group mean centered Xij
    Xij_gmc <- Xij-Xj


# RANDOM EFFECTS
  eij <- rnorm(Ng*Nj, 0, sqrt(lvl1unexp)) #individual level variance a
round the group reg line
  u0j <- rep(rnorm(Ng, 0, sqrt(lvl2unexp)), each = Nj) #random group m
ean variance around the grand mean


# OUTCOME VAR
  #Level 2 equations
    B0 <- gamma00 + u0j
    Bw <- gamma10w
    Bb <- gamma01b
  #Level 1 equation (Note: with values inserted it becames the combine
d model)
    y = B0 + Bw*(Xij-Xj) + Bb*(Xj) + eij
    yj <- rep(aggregate(y, list(grpid), mean)[, 2], each = Nj) # group
```

```
 means of Y
    any.data <- as.data.frame(cbind(grpid, y, Xij, Xj, Xij_gmc = (Xij-
Xj), eij, u0j))
```

Case 3 – *Between-group regression coefficient close to zero*

```
# GROUP DETAILS
  Ng <-         10 #number of groups
  Nj <-         500 #number of members (try with larger samples)
# VARIANCEs
  Xj_var <-     1 #muliplier term to make Xj variance big or small
  Xij_var <-    1 #sampling variance for Xij generation
  lvl1unexp <-  5
  lvl2unexp <-  5
# GAMMAs
  gamma00 <-    0 #Expected Y-grand mean (intercept for the average gr
oup)
  gamma10w <-   10 #Within group effect
  gamma01b <-   0 #Between group effect


# DATA GEN
  grpid <- rep(seq(1, Ng), each = Nj) # group ids
  #LVL-1 predictor: any continuos random variable (Xij)
    #Note: Xij mean depends on group membership
    Xj_gen <- seq(0, max(Xj_var*unique(grpid)), by = Xj_var) # vector
of group means from which to pick the Xij
    set.seed(2018)
    Xij <- vector("list", Ng)
      for(g in 1:max(grpid)){
        Xij[[g]] <- rnorm(Nj, Xj_gen[g], Xij_var)
      }
      Xij <- stack(as.data.frame(do.call("cbind", Xij)))[, 1]
  #LVL-2 predictor: Xij group means (Xj)
    Xj <- rep(aggregate(Xij, list(grpid), mean)[, 2], each = Nj)
  #Group mean centered Xij
    Xij_gmc <- Xij-Xj


# RANDOM EFFECTS
  eij <- rnorm(Ng*Nj, 0, sqrt(lvl1unexp)) #individual level variance a
round the group reg line
  u0j <- rep(rnorm(Ng, 0, sqrt(lvl2unexp)), each = Nj) #random group m
ean variance around the grand mean


# OUTCOME VAR
  #Level 2 equations
    B0 <- gamma00 + u0j
    Bw <- gamma10w
    Bb <- gamma01b
  #Level 1 equation (Note: with values inserted it becames the combine
d model)
```

```
    y = B0 + Bw*(Xij-Xj) + Bb*(Xj) + eij
    yj <- rep(aggregate(y, list(grpid), mean)[, 2], each = Nj) # group
 means of Y
    any.data <- as.data.frame(cbind(grpid, y, Xij, Xj, Xij_gmc = (Xij-
Xj), eij, u0j))
```

Case 4 – *Groups with similar means*

First, the parameters definition.

```
# GROUP DETAILS
  Ng <-          10 #number of groups
  Nj <-          500 #number of members (try with larger samples)
# VARIANCEs
  Xj_var <-      .01 #muliplier term to make Xj variance big or small
  Xij_var <-     1 #sampling variance for Xij generation
  lvl1unexp <-   5
  lvl2unexp <-   10
# GAMMAs
  gamma00 <-     0 #Expected Y-grand mean (intercept for the average gr
oup)
  gamma10w <-    3 #Within group effect
  gamma01b <-    0 #Between group effect
```

Subsequently, the data generation code with a different command to create the group means `Xj_gen`. In the previous examples the group means used to generate the group observations were the 10 integers included in the interval 0-9. In the present case, these means are now sampled from a normal distribution with mean 0 and standard deviation (`Xj_var`) 1.

```
# DATA GEN
  grpid <- rep(seq(1, Ng), each = Nj) # group ids
  #LVL-1 predictor: any continuos random variable (Xij)
    #Note: Xij mean depends on group membership
    set.seed(180524)
    Xj_gen <- rnorm(Ng, 0, Xj_var)
    Xij <- vector("list", Ng)
      for(g in 1:max(grpid)){
        Xij[[g]] <- rnorm(Nj, Xj_gen[g], Xij_var)
      }
    Xij <- stack(as.data.frame(do.call("cbind", Xij)))[, 1]
  #LVL-2 predictor: Xij group means (Xj)
    Xj <- rep(aggregate(Xij, list(grpid), mean)[, 2], each = Nj)
  #Group mean centered Xij
    Xij_gmc <- Xij-Xj
# RANDOM EFFECTS
  eij <- rnorm(Ng*Nj, 0, sqrt(lvl1unexp)) #individual level variance a
round the group reg line
  u0j <- rep(rnorm(Ng, 0, sqrt(lvl2unexp)), each = Nj) #random group m
```

```
ean variance around the grand mean


# OUTCOME VAR
  #Level 2 equations
    B0 <- gamma00 + u0j
    Bw <- gamma10w
    Bb <- gamma01b
  #Level 1 equation (Note: with values inserted it becames the combine
d model)
    y = B0 + Bw*(Xij-Xj) + Bb*(Xj) + eij
    yj <- rep(aggregate(y, list(grpid), mean)[, 2], each = Nj) # group
 means of Y
    any.data <- as.data.frame(cbind(grpid, y, Xij, Xj, Xij_gmc = (Xij-
Xj), eij, u0j))
```

Case 5 – *Low explanatory power for the predictor group means*

As in the previous cases, the first step is to define the parameters used to generate the data. The main difference is that a variance covariance matrix is created to be later used in sampling the values of X and y from a multivariate normal distribution.

```
# GROUP DETAILS
Ng <-          10 #number of groups
Nj <-          500 #number of members (try with larger samples)
# VARIANCEs
Xj_var <-      1 #variance for the noise added to the generating group
means
gamma10w <-    .7 #Within group effect (as correlation coefficient 0-1)

# VAR-COVAR matrix for multivariate normal distribution
q = sqrt(10)  # intercepts standard deviation^2 (var(u0j))
s = sqrt(10)  # if 0, Y does not depend on X
r = gamma10w  # how strongly Y depends on X correlation between slopes
 and intercept
cov.matrix <- matrix(c(q^2, r*q*s,
                       r*q*s, s^2),
                     nrow = 2,
                     byrow = TRUE)
```

Then, the group means of Y and X are defined. The means will be used to sample the individual cases in each group. Two means of Y are selected: 5 and 15. Five out of the 10 total groups will have a Y-mean of 5 and the other five a mean of 15. For what concerns the means of X, five values are selected (0, 10, 20, 30, and 40) and groups will be assigned each value in couples (i.e. two groups will have mean value of X corresponding to 0, two group will have 10, etc.)

```
Xj_gen <- rep(seq(0, 40, by = 40/4), each = 2)
Yj_gen <- rep(c(5, 15), Ng/2)
```

Subsequently, noise is added to the group means of X so as to create some variance in them.

```
set.seed(180524)
Xj_noise <- rnorm(Ng, 0, sqrt(Xj_var))
Xj_gen <- Xj_gen + Xj_noise
```

Next, the individual values of Y and X are sampled from a multivariate normal distribution.

```
datalist <- vector("list", Ng)
for(i in 1:Ng){
  set.seed(4040)
  data <- rmvnorm(Nj, mean = c(Xj_gen[i], Yj_gen[i]), sigma = cov.matr
ix)
  data <- as.data.frame(data)
  data$grpid <- c(rep(i, Nj))
  datalist[[i]] <- data
}
any.data <- do.call("rbind", datalist) # putting togheter the observat
ions into one data set
any.data <- data.frame(grpid = any.data$grpid,
                       y = any.data$V2,
                       Xij = any.data$V1)
```

Finally, the group mean level-2 variable and the group-mean centered version of Xij are added to the dataset.

```
any.data <- data.frame(grpid = any.data$grpid,
                       y = any.data$y,
                       Xij = any.data$Xij,
                       Xj = rep(aggregate(any.data$y, list(any.data$gr
pid), mean)[, 2], each = Nj),
                       Xij_gmc = any.data$Xij - rep(aggregate(any.data
$y, list(any.data$grpid), mean)[, 2], each = Nj))
```

Case 6 – *Random Intercept and Random Slope Models*

The generation process follows the model strategy used for the first four cases. However, some adjustments to the code are required to introduce random slopes. First, the parameters are defined:

```
# GROUP DETAILS ####
    Ng <- 10 #number of groups
    Nj <- 500 #number of members
    grpid <- rep(seq(1, Ng), each = Nj)
  # UNEXPLAINED VARIANCE
    lvl1unexp <- 3
    lvl2unexp <- 5
    tau01 <- 5 #random slopes variance
```

```
  #Gammas
    gamma00 <- 0 #Expected Y-grand mean (intercept for the average gro
up)
    gamma10w <- 5 #Within group effect
    gamma01b <- 2 #Between group effect
```

`tau01` is a new parameter that was not defined in previous cases. It is the variance of the random slopes. Next, individual observations for the predictor are sampled. In each group, the observations are sampled from a normal distribution with standard deviation 1 and mean corresponding to the id of the group membership. The level-2 predictor is formed by computing the means of the sampled `Xij` in each group and assigning it to each individual. Finally, the group-mean centered version of X, `Xij_gmc,` is computed.

```
# PREDICTORS ####
#LVL-1 predictor: any continuos random variable (Xij)
  #Note: Xij mean depends on group membership
  Xij <- vector("list", Ng)
  for(g in 1:max(grpid)){
    Xij[[g]] <- rnorm(Nj, unique(grpid)[g], 1)
  }
  Xij <- stack(as.data.frame(do.call("cbind", Xij)))[, 1]
#LVL-2 predictor: Xij group means (Xj)
  Xj <- rep(aggregate(Xij, list(grpid), mean)[, 2], each = Nj)
#Group mean centered Xij
  Xij_gmc <- Xij-Xj
```

The following step is the sampling of the random effects. First, the individual deviation scores are sampled from a normal distribution with mean 0 and variance `lvl1unexp`, defined above. Next, the random intercepts and slopes are sampled from a multivariate normal distribution with variance-covariance matrix (`cov.matrix`) and means 0. The variance used in the variance-covariance matrix are defined at the beginning of the code (`lvl2unexp`, and `tau01`). After creating a data frame of random effects (`random.effects`), they are extracted by selecting the proper column.

```
#Random effects
  eij <- rnorm(Ng*Nj, 0, sqrt(lvl1unexp)) #individual level variance a
round the group reg line
  q = sqrt(lvl2unexp)  #intercepts standard deviation^2 (var(u0j))
  s = sqrt(tau01)     #slopes standard deviation^2 (var(u1j))
  r = .9             #correlation between slopes and intercept
  cov.matrix <- matrix(c(q^2, r*q*s, r*q*s, s^2),
                       nrow = 2,
                       byrow = TRUE)
  random.effects <- rmvnorm(Ng, mean = c(0, 0), sigma = cov.matrix)
```

```
  u0j <- rep(random.effects[, 1], each = Nj)
  u1j <- rep(random.effects[, 2], each = Nj)
```

Finally, the Y variable is computed, and the dataset is created by joining all variables in a data frame.

```
#Level 2 equations
  B0 <- gamma00 + u0j
  Bw <- gamma10w + u1j
  Bb <- gamma01b
#Level 1 equation (Note: with values inserted it becames the combined
model)
  y = B0 + Bw*(Xij-Xj) + Bb*(Xj) + eij
  yj <- rep(aggregate(y, list(grpid), mean)[, 2], each = Nj) # group m
eans of Y
  any.data <- as.data.frame(cbind(grpid, y, Xij, Xj, Xij_gmc, eij, u0j
, u1j))
```

Learning Process Reflection

The process of studying and writing the First Year Paper has taught me a lot of things. First of all, I have delved deep into Multilevel Analysis learning much more about how this statistical tool can be complicated in its interpretation but extremely powerful.

I have learned how to try to define a problem by comparing findings coming from different sources in the literature. Mixed effects models are used in many different disciplines, each of which often has specific takes on it, with its own notation, applications and costumes. Finding relevant articles, combining and understanding insights coming from a diverse range of disciplines has been a challenge at first, and a resource at last.

I have learned how to propose an alternative statistic, by explaining the root of the idea, and its advantages and disadvantages over the more established measures in the literature.

The hardest part has been generating datasets on which to test the performances of the measures. The challenge was twofold: on the one hand, it has been conceptually hard to come up with scenarios that stressed the measures in the desired ways, to expose their strengths and weaknesses; on the other hand, I had to learn how to use R to achieve what I wanted. Making fictitious datasets plausibly reflect real situations was also part of the challenge.

Apart from the technical aspects, I also had to learn how to manage my time and efforts in an efficient way. This has actually been more difficult than I expected. I had to learn how to pace the work relying on my judgment instead of external constraints and the results have not always been optimal. However, I believe that this experience has improved my abilities in this regard.