RUNNING HEAD: ICC and Dependency

Group Dependency in the Presence of Small Intraclass Correlation Coefficients: An Argument in Favor of Not Interpreting the ICC.

J. Kyle Roberts
Baylor College of Medicine

Paper presented at the annual meeting of the American Educational Research Association, April 10, 2007. Correspondence should be sent to J. Kyle Roberts, One Baylor Plaza, MS: BCM 411, Houston, TX 77030-3411, or jkrobert@bcm.edu. Funding for the development of this manuscript was provided by Grant Number R25 ES10698 from that National Institutes of Environmental Health Sciences division of the National Institutes of Health.

Group Dependency in the Presence of Small Intraclass Correlation Coefficients: An Argument in Favor of Not Interpreting the ICC.

In multilevel analysis, intraclass correlation (ICC) is the proportion of total variance that is between groups. In a two-level model, the ICC is found by dividing the variance at the highest level (in this case the level-2) by the sum of the variances at the lowest level and the highest level. In other words, as Equation 1 explains, ICC (ρ) for a two-level model is the proportion of group level variance from the total variance, where σ_{u0}^2 represents the level-2 variance and σ_{e0}^2 represents the level-1 variance:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2} \ . \qquad (1)$$

Hox (2002) explains the ICC as the "proportion of the variance explained by the grouping structure in the population" (p. 15). Kreft & de Leeuw (1998) describe the ICC as "the degree to which individuals share common experiences due to closeness in space and/or time" (p. 9). This concept is important to the researcher because if intraclass correlation exists, Kreft & de Leeuw argue (1998, p. 9), then the traditional linear model must be abandoned because the assumption of independent observations has been violated

It is helpful here to illustrate the importance of ICC. Let us suppose that a researcher has collected data on science achievement from four schools where one is urban, one is suburban, one is private, and one is rural. The traditional OLS model would assume that each of these observations was independent of the context/school in which the data were collected, therefore neglecting intraclass correlation. Thus, the prediction of student scores in science achievement would be estimated irrespective of the type of school that the student attended. This could be an erroneous assumption, because as all researchers realize, a 4.0 grade point average (GPA) at a highly selective high achieving preparatory school is not the same as a 4.0 GPA at a low performing urban school.

The problem with the above argument by Kreft & de Leeuw is that it is not reciprocal. In the presence of a small ICC, it does not stand to reason that the assumption of independence of observations has not been violated. The purpose of this paper is to show that although a researcher might assume no group dependence in the presence of a small ICC computed for a null (fully unconditional) multilevel ANOVA, the degree of observational dependence actually is determined by the nature of the covariates/predictors chosen to be included in the model.

Some might assume that researchers competent to understand applications of HLM would not fall into the trap of assuming that the reciprocal of the Kreft and de Leeuw (1998, p. 15) statement is also true. However, this type of thinking is commonly illustrated in passages like the following:

> Determining the proportion of the total variance that lies systematically between schools, called the intraclass correlation (ICC), constitutes the first step in an HLM analysis. We conduct this analysis with a fully unconditional model, which means that no student or school characteristics are considered. This first step can also indicate whether HLM is needed or whether a single level analytic method is appropriate. *Only when the ICC is more than trivial* (i.e., greater than 10% of the total variance in the outcome) *would the analyst need to consider multilevel methods* [emphasis mine]. Ignoring this step (i.e., assuming an ICC of either 0 or 1) would be inappropriate if the research question were multilevel. Investigation of contextual effects, I argue, is by nature a multilevel question. (Lee, 2000, p. 128).

Statements like this are somewhat troubling in that they wrongly put forward the idea that multilevel analysis is only appropriate when the ICC is greater than .10. The current paper illustrates cases where the inclusion of certain predictor variables into the model could "create" group dependence in the presence of a small ICC from the null model. This will be illustrated with two different types of models. The first model is an instance where the homogeneity of variance assumption (Raudenbush & Bryk, 2002, p. 255) is not met, thus indicating that the Level-1 errors are not independent and identically normally distributed (IID). Even though this type of model would be a violation of our assumptions for a multilevel analysis, this type of data is illustrated to show the importance of checking these assumptions and the potential pitfalls that can come about from not checking them. The second model is an instance where the homogeneity of variance assumption is met.

An example where there is no homogeneity of variance in the dependent variable.

A simple simulation was constructed in the R programming environment to illustrate the problem of small ICC's in multilevel analysis. In this simulation, data for 30 groups was created with 30 individuals in each group for a total of 900 observations. In both this example, and in the following example where the homogeneity of variance assumption is met, two models will be illustrated. The first model, or the null model is defined as:

$$dv_{ij} = \gamma_{00} + u_{0j} + e_{ij} \ , \qquad (2)$$

with σ_e^2 Level-1 variance and σ_{u0}^2 Level-2 variance. The full model is defined as:

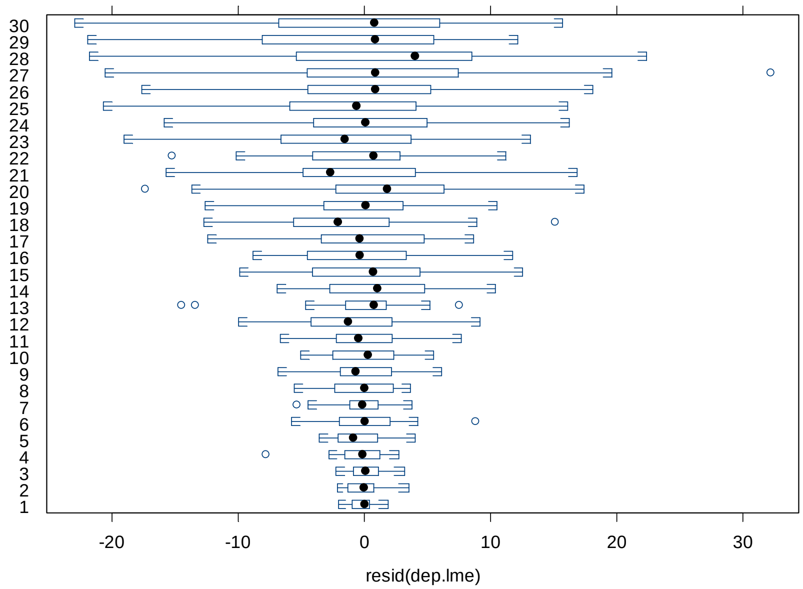
$$dv_{ij} = \gamma_{00} + \gamma_{10}(iv_{ij}) + u_{0j} + u_{1j}(iv_{ij}) + e_{ij} \ , \qquad (3)$$

with σ_e^2 Level-1 variance, σ_{u0}^2 variance for the intercept, and σ_{u1}^2 variance for the slope of iv_{ij}. Typically, we would assume e_{ij} ~ independently N(0, σ_e^2), and both u_{oj} ~ independently N(0,

σ^2_{u0}) and u_{1j} ~ independently N(0, σ^2_{u1}), but in the first illustration, we are not following these assumptions. These assumptions are held, however, in the second illustration.

Figure 1 shows box and whisker plots for the residual scores for the null model from Equation 2.

Figure 1 – Residual plots for the fitted null model



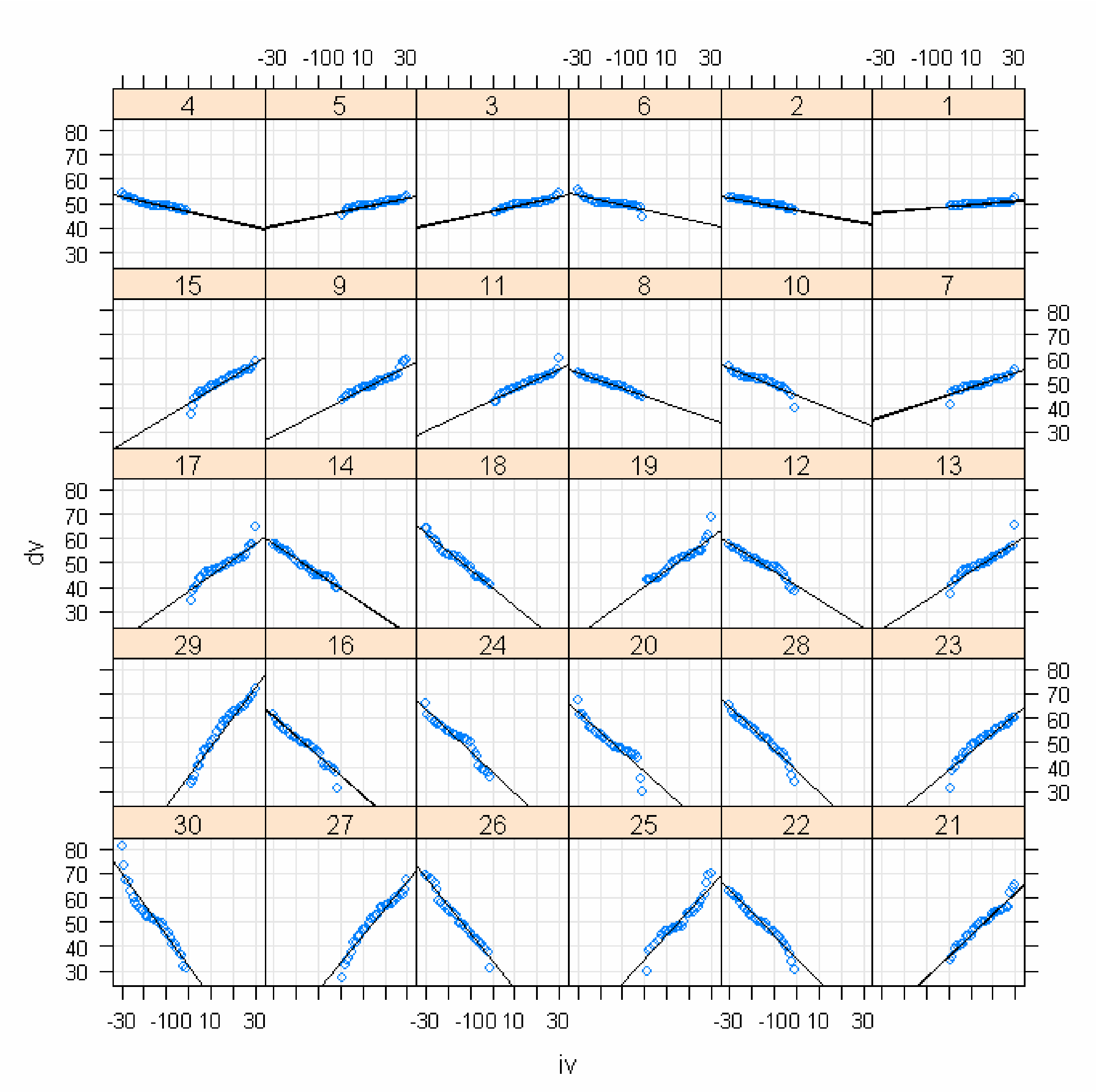
As can be seen from the above graph, each of the group's mean is very close to the grand mean. However, it is obvious that this model violates the assumption that the Level-1 error terms are IID. As will be illustrated, neglecting to check this assumption could lead to errors associated with the ability to draw conclusions about group dependence from the ICC. Table 1 presents the results from running the unconditional model (Equation 2) and the full model with one predictor and adding the random effect for this predictor (Equation 3).

Table 1.

	M_0: Null model		M_1: + dv & random	
Fixed Effects:	estimate	s.e.	estimate	s.e.
Intercept γ_{00}	49.89	0.21	40.56	0.91
Slope γ_{10}			-0.01	0.13
Random Effects:				
Level-1 effect, σ^2_{e0}	38.02		2.48	
Intercept, σ^2_{u0}	< .00001		24.23	
Slope, σ^2_{u1}			0.48	
$COV(u_0, u_1)$,			0.02	
Fit:				
X^2	5828.70		3724.25	
AIC	5834.70		3736.25	
BIC	5849.11		3765.05	

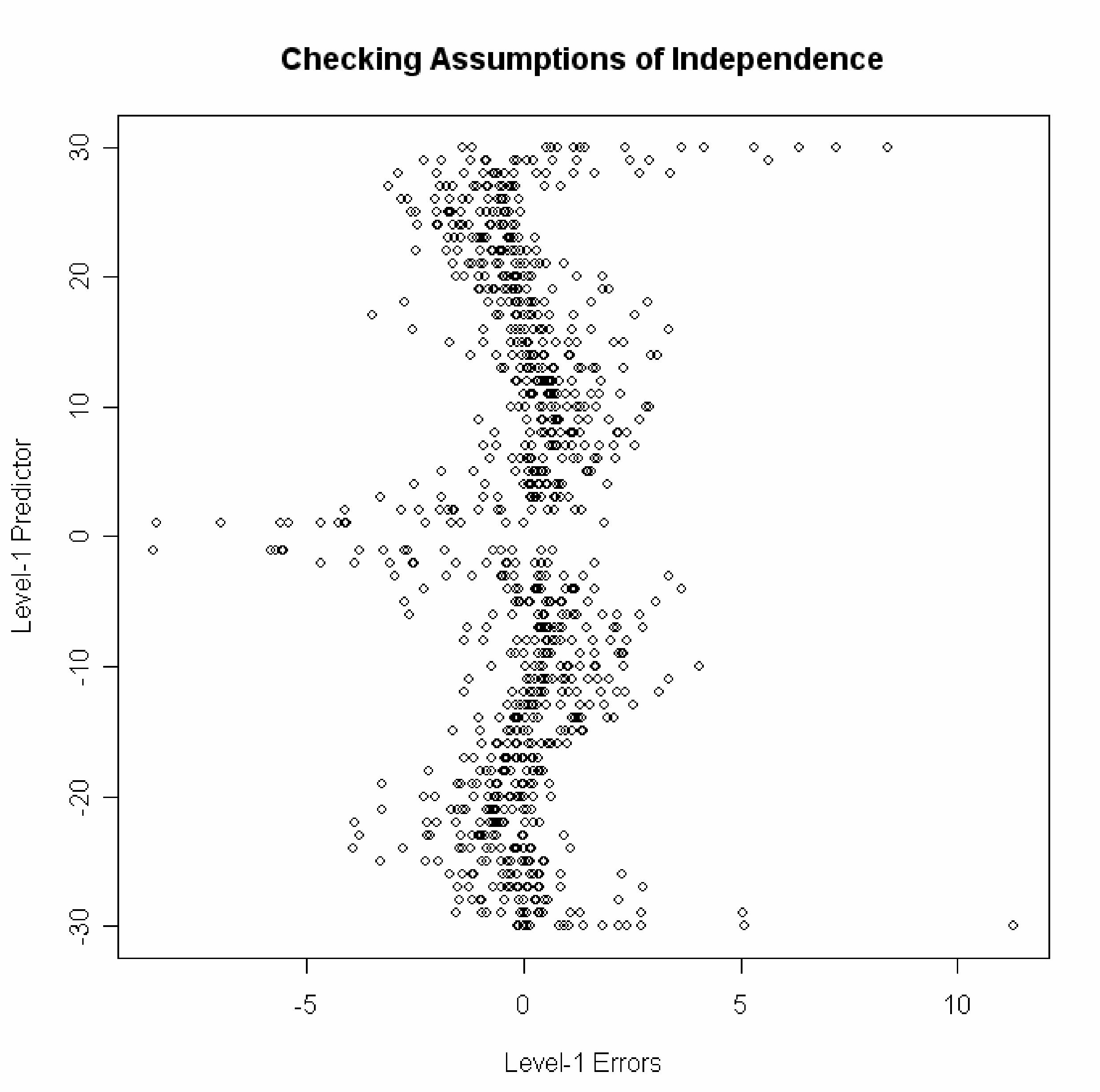
As can be seen from Table 1, the ICC is so small that it is barely worth computing. However, a wise researcher would probably look at the null model residuals plots in Figure 1 with some speculation and wonder whether or not the groups are indeed identical in reference to their relationship with the dependent variable. However, when viewing the result from model M1, we can see that a large amount of variability suddenly “appears” among the intercepts. Viewing the prediction lines for each group based on M1 from Table 1 gives some indication as to the reason for these results. These results are seen in Figure 2.

Figure 2 – Prediction lines based on model M1 for all 30 groups



What also "appears" from model M1 is the fact that the Level-1 predictor is not independent of the Level-1 error terms. In an investigation of this assumption we can see from the plots in Figure 3 that there is indeed a relationship between the error term and the predictor "iv."

Figure 3 – Plot of the Relationship Between the Level-1 Errors and the Independent Variable

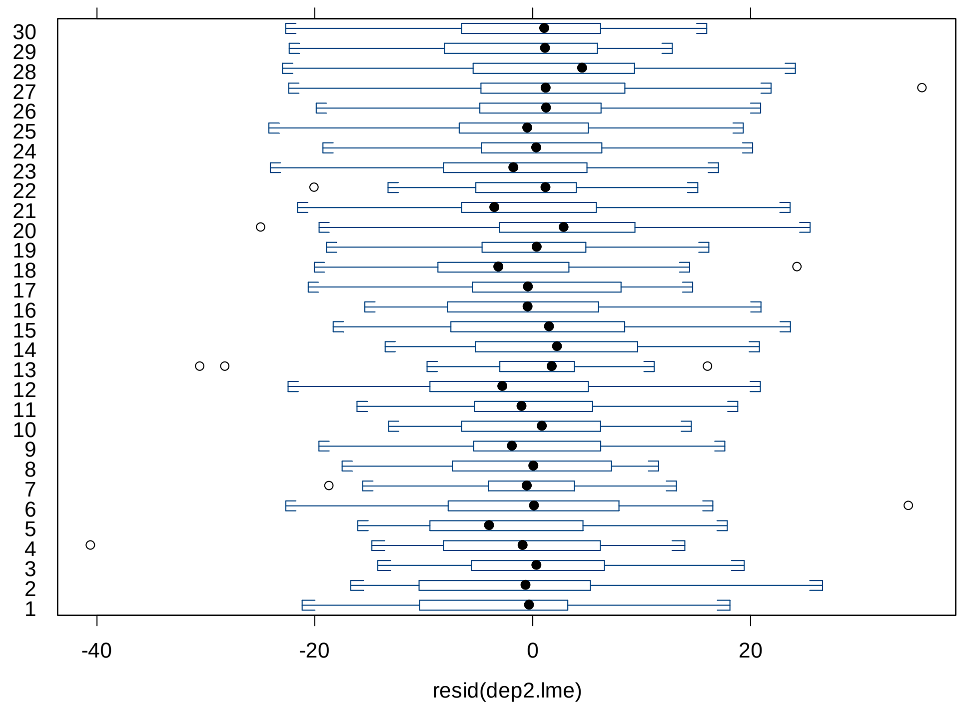


It is easy to see from this dataset, that violating just a few of the assumption of the multilevel model can produce this masked group dependency. Although it might be assumed

An example where there IS homogeneity of variance in the dependent variable.

An argument could be made that the above example is a special case where a homogeneity of variance assumption might be grossly violated and that this special case is the only time when a researcher need be wary of interpreting the ICC as a measure of group dependence. In order to counter this argument, a set of simulated data was again created in which all of the Level-2 groupings have homogeneity of variance. The plot of the residuals from this model can be seen in Graph 4.

Figure 4 – Residual plots for the fitted null model



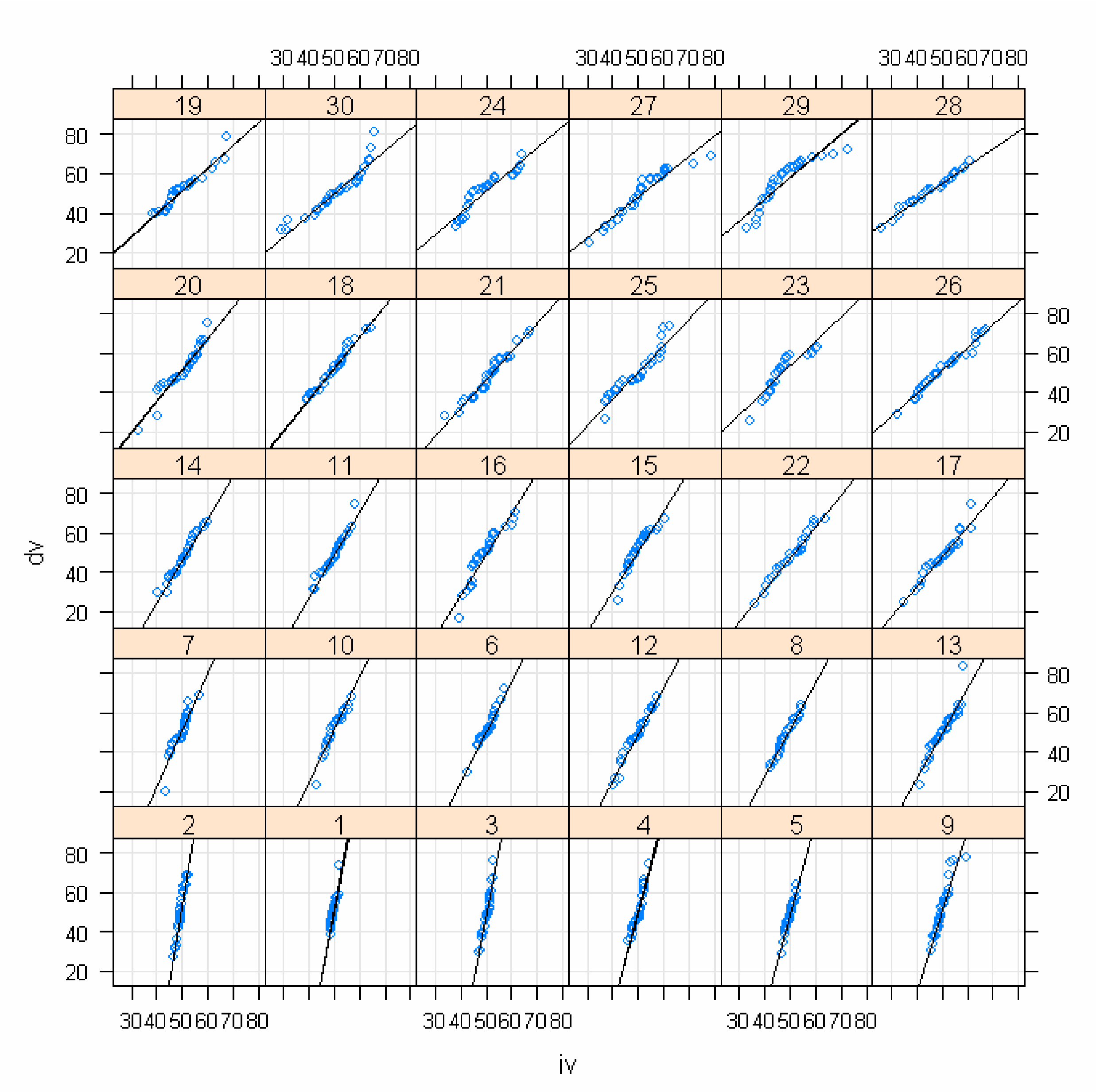
The results from the fitted models based on this data can be seen below in Table 2.

Table 2 – Model fit estimates for data with homogeneity of variance.

	M_0: Null model		M_1: + dv & random	
Fixed Effects:	estimate	s.e.	estimate	s.e.
Intercept γ_{00}	49.61	0.33	-74.32	18.16
Slope γ_{10}			2.47	0.36
Random Effects:				
Level-1 effect, σ^2_{e0}	98.29		8.01	
Intercept, σ^2_{u0}	> .00001		9821.94	
Slope, σ^2_{u1}			3.87	
$COV(u_0, u_1)$,			-195.04	
Fit:				
X^2	6682.62		4683.29	
AIC	6688.62		4695.29	
BIC	6703.03		4724.09	

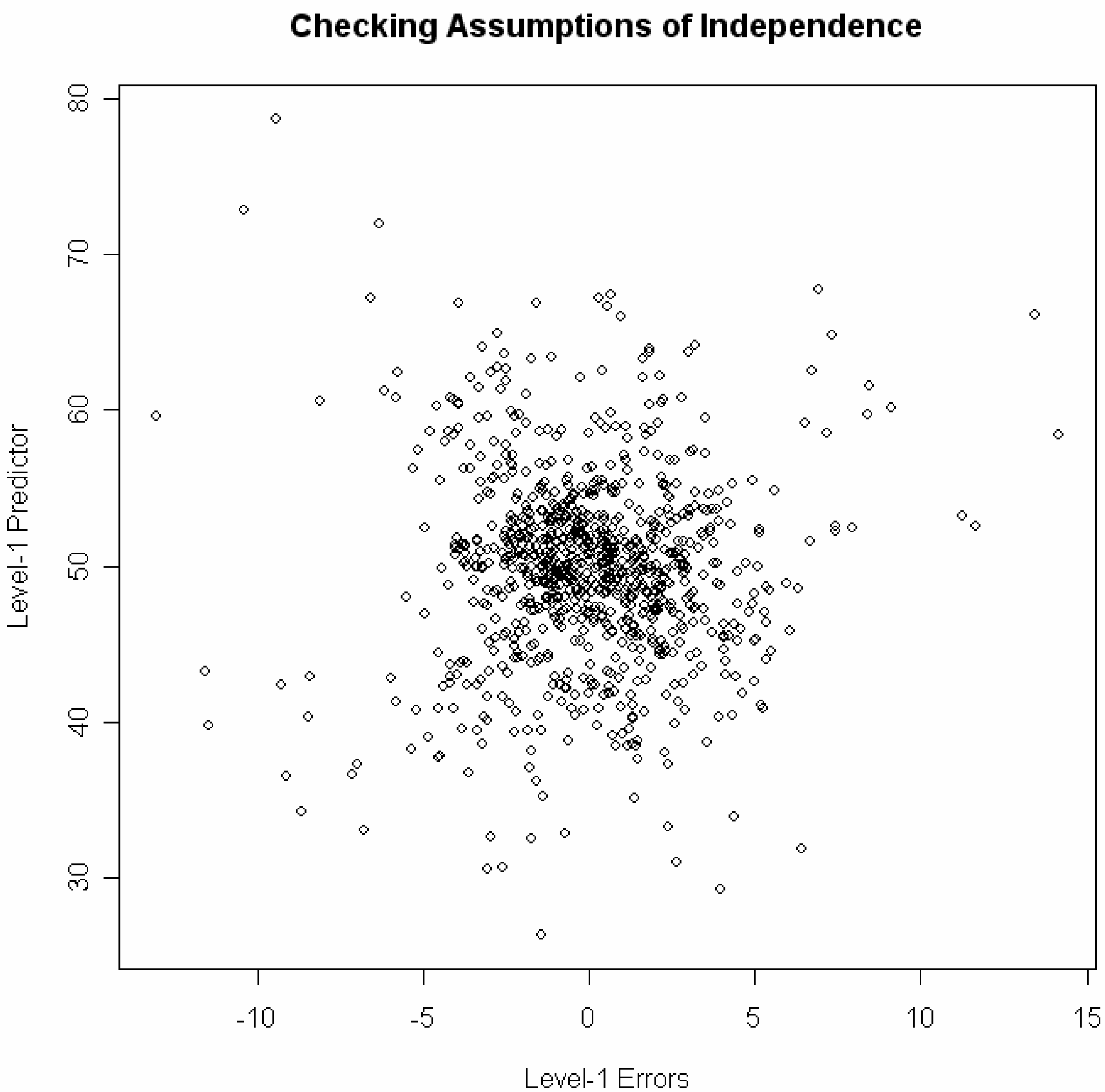
Just like in the first example, the ICC from this null model is amazingly small. However, when we add in the predictor and the random effect, we see a large amount of variation among the groups. Further evidence of this can be seen in Figure 5 of the fitted prediction lines based on M1.

Figure 5 – Prediction lines based on model M1 for all 30 groups



This case is equally as dramatic as the previous case when considering the role of group dependence. And unlike as was the case in the model from the previous section, this model meets both our assumption of IID and the Level-1 predictor being independent of Level-1 errors (see Figure 6).

Figure 6 – Plot of the Relationship Between the Level-1 Errors and the Independent Variable



As the "iv" x Level-1 error graph shows (Figure 6), it is reasonable to say that we met the assumption of independence. As a result, it seems necessary to discuss what is going on in this dataset to produce the results described in Table 2. In this example, the dependent variable seemed to behave similarly across all groups. But when the independent variable was added into the model, the relationship between the dependent variable and independent variable changed drastically across groups.

As was discussed previously, student grade point average (gpa) provides a perfect illustration of this effect. Suppose that we were monitoring 30 high schools and had the gpa from 30 randomly selected students from each of those 30 schools. It might be reasonable to assume that the 30 randomly selected students in each school would have a mean gpa ~3.0 and would be normally distributed within that school with some variance, σ^2. This might be a safe assumption since regardless of the difficulty of the school, it stands to reason that within that school, these properties of gpa would hold true.

Suppose, also, in this example that we used scores on the SAT college entrance exam to predict high school gpa. Although this would seem somewhat backwards, it still illustrates a case where the independent variable now creates group dependence. In this case, the relationship between gpa and SAT, although still linear and most likely having a similar strength of relationship (in terms of a single R^2-type metric for each school) across schools, would have differing magnitudes of slope (or rate of change) in different schools.

For example, in Figure 5 consider groups 7 and 28. In this example, a small change in "iv" produces a small change in "dv" for group 28, but a large change in "dv" for group 7. This is similar to our gpa example in that in the gpa example we may have samples schools with drastically different achievement levels which the dependent variable gpa does not capture, but which SAT does capture.

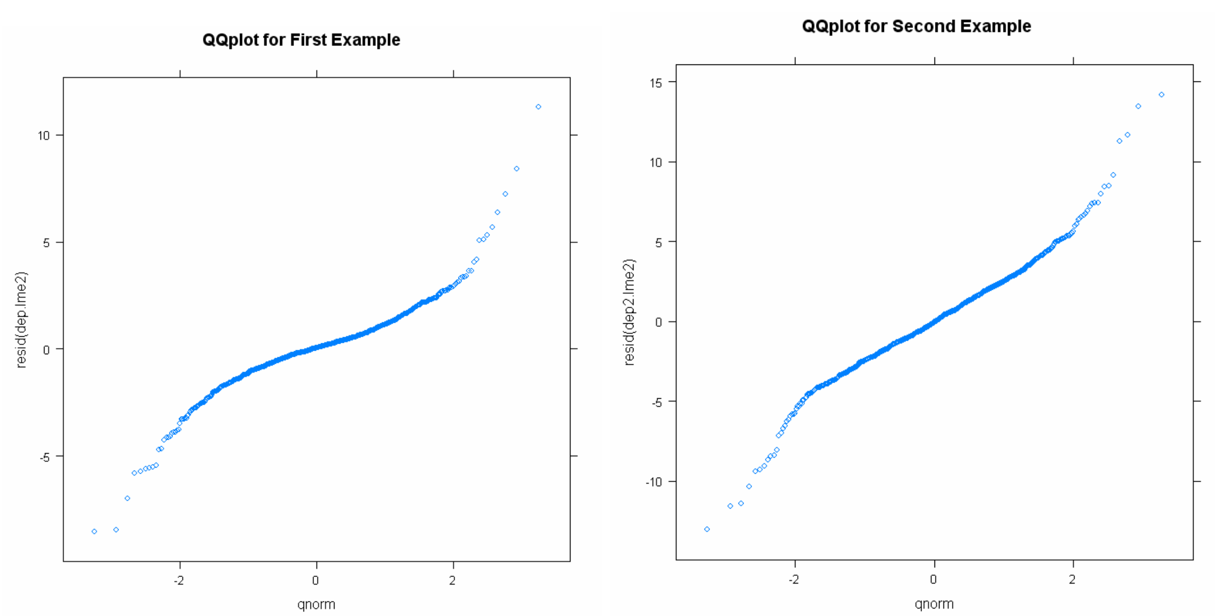
Discussion.

There is one gaping problem when considering the above two illustrations of datasets where group dependence is "created." Snijders and Bosker (1999) say that "for reasonably large datasets, a decrease by magnitude [of the explained variance of the intercept] of 0.05 or more should be taken as a warning of possible misspecification” (p. 123). As can be seen from both Table 1 and Table 2, the level-2 explained variance for both intercepts is an incredibly large *negative* number when computed based on the formula:

$$R_2^2 = \frac{\sigma_{u0|b}^2 - \sigma_{u0|m}^2}{\sigma_{u0|b}^2} \quad , \qquad (4)$$

where $|b$ and $|m$ represent the baseline and full models, respectively, and σ_{u0}^2 represents the Level-2 variance. While these values are much larger than 0.05, it still stands to reason for the second data example, that we have reasonably managed to meet these assumptions. The one assumption that could still prove a problem, however, is whether or not the Level-1 errors are independent and identically normally distributed (IID). Figure 7 shows us the plots for Model M1 in both the First and Second Example from Tables 1 and 2.

Figure 7 – QQPlot for the Level-1 Residuals in Both Examples



It is easy to see from Figure 7 that the first example violates this assumption of IID. The second example is a little more difficult to ascertain whether or not the plots follow a strict 45-

degree plot, and probably a little more investigation would be needed, but it might could be argued that this example meets the assumptions of a multilevel model.

These examples, although heuristic in nature, do give several cautions to the researcher in multilevel modeling. The first, from the first illustration, is that a failure to check distributional assumptions could lead to a multilevel model in which group dependence is masked, but only appears after multiple assumptions have been violated.

The second caution is that there may never be a time when it is acceptable to say that the only time that multilevel analysis is appropriate is when ICC is beyond some threshold. The second example showed clearly that even though the ICC was near zero for the null model, group dependence may still exist depending on the nature of the covariates introduced into the model.

The final caution is to consider that even though we may be quick to assert that naturally occurring groups in our data behave similarly, we cannot fully assert that they are indeed similar with respect to all cases because there may be covariates which we have not discovered that "create" group heterogeneity.

References

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahway, NJ: Erlbaum

Kreft, I. & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Lee. V. E. (2000). Using Hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, *35*(2), 125-141.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks, CA: Sage.

R Code for Simulating Datasets

```
####################################################################################
#### Solution in which the homogeneity of variance assumption is not met
set.seed(100)
myRands <- lapply((lapply(seq(1, 10, length = 30), rnorm, n = 30, mean = 50)), sort)
depend <- data.frame(index = rep(1:length(myRands), each = 30), dv = unlist(myRands),
       iv = rep(c(1:30, -1:-30), 15))

#### Run the null model
dep.lme<-lmer(dv ~ 1 + (1|index), depend)

#### Get residual box and whiskers for null model
bwplot(index~resid(dep.lme), depend)
bwplot(index~dv, depend)

#### Run full model
dep.lme2<-lmer(dv ~ iv + (iv|index), depend)

#### Get residual box and whiskers for full model
bwplot(index~resid(dep.lme2), depend)

#### Graphing
with(dep.lme2, {
    cc <- coef(.)$index
    xyplot(dv ~ iv | index,
                 index.cond = function(x, y) coef(lm(y ~ x))[1],
                 panel = function(x, y, groups, subscripts, ...) {
                     panel.grid(h = -1, v = -1)
                     panel.points(x, y, ...)
                     subj <- as.character(index[subscripts][1])
                     panel.abline(cc[subj,1], cc[subj, 2])
                     panel.lmline(x, y, ...)
                 })
})

#### Assumptions
plot(depend$iv~residuals(dep.lme2), xlab="Level-1 Errors", ylab="Level-1 Predictor",
main="Checking Assumptions of Independence")

qqmath(~resid(dep.lme2))
```

```
###########################################################################################
#### Solution in which the homogeneity of variance assumption is met

#### Generate the dataset
set.seed(100)
myRands2 <- lapply((lapply(rep(10, 30), rnorm, n=30, mean=50)), sort)
myIV2 <- lapply((lapply(seq(1, 10, length = 30), rnorm, n = 30, mean = 50)), sort)
depend2 <-data.frame(index=rep(1:length(myRands2), each=30), dv = unlist(myRands2),
       iv = unlist(myIV2))

#### Run the null model
dep2.lme<-lmer(dv ~ 1 + (1|index), depend2)

#### Get residual box and whiskers for null model
bwplot(index~resid(dep2.lme), depend2)

#### Run full model
dep2.lme2<-lmer(dv ~ iv + (iv|index), depend2)

#### Get residual box and whiskers for full model
bwplot(index~resid(dep2.lme2), depend2)

#### Graphing
with(dep2.lme2, {
    cc <- coef(.)$index
    xyplot(dv ~ iv | index,
                 index.cond = function(x, y) coef(lm(y ~ x))[1],
                 panel = function(x, y, groups, subscripts, ...) {
                     panel.grid(h = -1, v = -1)
                     panel.points(x, y, ...)
                     subj <- as.character(index[subscripts][1])
                     panel.abline(cc[subj,1], cc[subj, 2])
                     panel.lmline(x, y, ...)
                 })
})

#### Assumptions
plot(depend2$iv~residuals(dep2.lme2), xlab="Level-1 Errors", ylab="Level-1 Predictor",
main="Checking Assumptions of Independence")

qqmath(~resid(dep2.lme2))
```