Paper 201-2010

Fundamental Diagnostics for Two-Level Mixed Models: The SAS® Macro MIXED_DX

Bethany A. Bell, Jason A. Schoeneberger, Grant B. Morgan
University of South Carolina
Jeffrey D. Kromrey, John M. Ferron
University of South Florida

ABSTRACT

Multilevel modeling has become a common analytic technique across a variety of disciplines including medicine and the social and behavioral sciences. However, because many researchers who use multilevel modeling in their research do not report if the data were screened for potential violations of distributional assumptions and outliers, it is unclear if these diagnostic procedures are being conducted. Thus, in an effort to make the process of checking the assumptions for multilevel models easier for the applied researcher, this paper provides a SAS macro for conducting two-level linear model diagnostics, including examinations of residual normality, linearity, homogeneity of variance, and influential outliers. By utilizing data from PROC MIXED ODS tables, the macro produces box-and-whisker plots, summary tables of statistical output, histograms, and plots of residuals by predicted values. Macro outputs are produced for both level-1 and level-2 data. This paper provides a discussion about the distributional assumptions of mixed models, an example of the macro language, and results from an executed example of the macro. Information for downloading the complete macro is also provided.

Keywords: MULTILEVEL MODELING, STATISTICAL ASSUMPTIONS, INFLUENCE DIAGNOSTICS, PROC MIXED, SAS MACRO

INTRODUCTION

As multilevel modeling techniques (also called hierarchical linear modeling or mixed modeling) continue to be used with increasing frequency across a variety of disciplines, it is important for researchers to conduct appropriate model diagnostics when using these statistical methods. Like most parametric statistical procedures, there are certain distributional assumptions underlying the validity of the Type I error control when conducting multilevel analyses. Thus, just as researchers were trained to evaluate the underlying assumptions for multiple regression and ANOVA, they, too, need to assess and report on the assumptions associated with multilevel models. Although most of the assumptions associated with two-level linear models are similar to OLS model assumptions (i.e., residual normality, linearity, and homogeneity of variance), few articles in which multilevel modeling techniques are used contain this information, despite recommendations to be transparent when reporting multilevel modeling research processes and findings (Ferron, Hogarty, Dedrick, Hess, Niles, & Kromrey, 2008).

One likely reason for the omission of this information could be related to the cumbersome nature of conducting model diagnostics of two-level linear models. First, it is not easy to locate all of the data necessary to conduct diagnostics for both level-1 and level-2 residuals. For example, the level-1 and level-2 residuals are stored in two different ODS tables, which not only have to be requested separately, but the commands necessary to request each ODS table occurs at two different places within the PROC MIXED code (i.e., the ODS table for level-1 residuals is an option that can be added to the model statement whereas the ODS table for level-2 residuals is requested through an ODS OUTPUT command). Second, the ODS table that contains the level-2 residuals does not contain predicted values for level-2 units. Thus, the data must be manipulated to generate these values. Third, although PROC MIXED does allow the INFLUENCE option on the model statement, the oodles of data produced by this option can be overwhelming, especially to a novice PROC MIXED user. Thus, in an effort to reduce the burden of conducting these important, yet unwieldy, analyses this paper provides a SAS macro to conduct diagnostic evaluations of two-level linear models, including examinations of residual normality, linearity, homogeneity of variance, and influential outliers. Diagnostics are conducted for both level-1 and level-2 data and the macro includes the data screening techniques recommended by authors such as Hox (2002), Longford (2001), and Raudenbush and Bryk (2002).

STATISTICAL ASSUMPTIONS AND THEIR EVALUATION

Raudenbush and Bryk (2002) correctly assert that the validity of inferences based on models depend on the degree to which assumptions are upheld about the structural and random parts of the model. They go on to suggest that “skillful analysts pay close attention to the assumptions required by their models” (p. 253). Unfortunately, based on reporting practices, it seems that the assumptions frequently go unexamined or under-examined by analysts conducting multilevel modeling (Ferron et al., 2008). However, it is not clear if this common omission is a function of a lack of understanding or the effort and time needed to assess model adequacy. This work seeks to address the latter issue by facilitating the generation of information necessary to assess to the tenability of the model assumptions.

Under Ordinary Least Squares (OLS) regression, the error terms are assumed to be normally distributed, independent, and homoscedastic. These same assumptions apply to linear multilevel models. In particular, the assumption of normally distributed errors must be made for both level-1 and level-2 variables, with violations adversely affecting level-2 estimated standard errors and inferential statistics (Raudenbush & Bryk, 2002). The effects of the violation at level-1 may include distorted random effect coefficients and variance-covariance components. To examine the distribution of the errors, data analysts should consider the following options. First, one might plot the standardized residuals against their normal scores to observe how closely the plot follows a diagonal line (Hox, 2002). Second, to assess the distribution of the overall residuals at levels one and two, one might also use histograms or box-and-whisker plots. Similarly, box-and-whisker plots of the level-1 residuals for each level-2 unit and of the level-2 residuals for each level-2 effect will help assess deviations from a normal distribution and identify extreme values. Third, to assess assumptions of normality, linearity, and homogeneity of variance simultaneously, analysts should examine the plot of predicted values against the residuals. This should be done for level-1 residuals, as well as the residuals for each level-2 effect. Scatterplots that contain roughly equivalent frequencies of points above and below their mean value, with no particular structure, provide evidence that the assumptions have not been violated (Hox, 2002).

To supplement the visual level-1 assumption diagnostics, summary statistics (i.e., skewness, kurtosis, variance and standard deviation, and a statistical test for normality such as the Shapiro-Wilk or Kolmogorov-Smirnov test) for level-1 residuals within each level-2 unit should also be reviewed for information regarding level-1 normality and variance homogeneity. In addition, Levene's test can be used to assess homogeneity of variance of the level-1 residuals across each level-2 unit. Next, when evaluating the tenability of assumptions for the level-2 residuals, multivariate summary statistics such as skewness, kurtosis, and Mahalanobis distances should be generated and reviewed. Multivariate normality can also be evaluated visually through a histogram of Mahalanobis distances. And, as with the level-1 residuals, Levene's test for assessing homogeneity of variance of the level-2 residuals can also be conducted for each level-2 effect.

The MIXED_DX macro provides analysts with a comprehensive approach to assessing the degree to which one or more of the linearity, normality, or homogeneity of variance assumptions has been violated. The output includes both visual (e.g., box-and-whisker plots, histograms, scatter plots) and statistical information for both level-1 and level-2 residuals. A unique feature of our macro is the calculation of level-2 predicted values for each level-2 effect. Unlike the level-1 ODS output table in SAS (generated through the outp option) that contains both residuals and predicted values for each level-1 observation, the level-2 ODS output table in SAS (SolutionR) does not contain predicted values. By calculating predicted values for each level-2 effect, our macro allows analysts to create and examine residual by predicted scatterplots for level-2 effects (something that cannot be done using the default SAS output alone). Furthermore, the MIXED_DX macro generates output for all level-2 units, with an option for a user-defined minimum of cases per level-2 unit to be included in these analyses.

INFLUENCE

Nearly all statistical analyses involve post-hoc diagnostic investigations to determine how well models fit the data. One of the more well-established diagnostics for ordinary least-squares linear regression is the detection of influential observations in determining parameter estimates. Previous work by Cook (1977), Belsley, Kuh, and Welsch (1980), and Andrews & Pregibon (1978) provided useful statistics for determining how much influence a particular observation had over estimated parameters or the performance of the model. With the development of more complex multilevel models, the detection of influential observations becomes more complex as well. For example, mixed model analyses generate more results than may be of potential interest to the researcher. A researcher may be interested in predicting a particular outcome variable, in which case fixed effect results are of primary interest, whereas another researcher might be interested in explaining the variance in a particular outcome, thus variance component results become the primary concern. Further, with multilevel models, the concept of influence is often applied to higher-level units (i.e., group or cluster influence) vs. with OLS diagnostics where influence diagnostics are concerned with the influence of individual observations. This interest in the influence of higher-level clusters or groups calls for set-deletion analyses where an entire cluster or group is removed to determine its influence vs. the removal of individual observations as is done in OLS diagnostic algorithms.

Littell, Milliken, Stroup, and Wolfinger (2006) recommend a “drill-down” approach to mixed model influence diagnostics, beginning with a global assessment of the influence on the overall model, followed by a more detailed exploration of the case-sets should they be warranted. The likelihood distance (Cook & Weisberg, 1982) provides an assessment of a unit's influence on the overall model. A group or unit's influence on parameter estimates can be determined using Cook's *D* or the multivariate DFFITS statistics. The larger the value for these statistics, the greater the influence a unit has on parameter estimates (i.e., the change in the parameter estimate is large relative to the variability of the estimate; Schabenberger, 2004). Researchers can also examine a unit's effect of the precision of an estimate through the covariance trace (COVTRACE) and covariance ratio (COVRATIO) statistics. For these statistics, benchmark values used to determine a unit's level of influence are zero and one, respectively (Schabenberger, 2004). Each of these influence diagnostics can be generated for both fixed effects and covariance parameters, however calculation for the latter requires an iterative process due to the potential impact of observations on covariance matrices. The INFLUENCE option within PROC MIXED allows the researcher to utilize non-iterative or iterative diagnostics, and provides the option to control the number of iterations when re-calculating estimates and covariance matrices (Littell et. al., 2006).

In addition to examining a unit's influence on the change in parameter estimates and the change in the precision of estimates, influence on fitted and predicted values can also be examined through the Predicted Residual Sum of Squares (PRESS) statistic. The PRESS provides a comparison of the predicted marginal mean and the observed mean when the predicted value is calculated without the deleted observation in question (Schabenberger, 2004). Again, larger values suggest more influential units. Other measures of overall influence provided through the INFLUENCE option in PROC MIXED include RMSE and Restricted Likelihood Distance (RLD). RMSE values represent the model RMSE with a specific level-2 unit deleted. On the other hand, RLD functions more as an 'indicator' statistic in that it does not convey what part of the model is likely to change given the removal of a specific level-2 unit. Instead, it suggests that the overall influence of a particular level-2 unit stands out comparatively to other level-2 units (Littell et al., 2006). To determine what model components are influenced by a given level-2 unit, analysts need to examine the individual influence statistics such as MDFFITS or COVRATIO.

To facilitate detection of influential observations in a two-level linear model, the MIXED_DX macro creates a ranked summary table of the influence statistics automatically created in the SAS ODS influence table (Influence). More specifically, for each statistic provided in the summary table, the macro employs the RANK procedure to determine the percentile rank of each statistic's value for each level-2 unit. MIXED_DX allows the user to choose a threshold percentile rank of interest for the detection of influential observations via the 'PR = ' argument to the macro. The default, 'PR = 90', will flag any level-2 unit with a percentile rank greater than 90 for each statistic included in the ODS output. To further facilitate inspection of the summary table, particularly for data sets with a large number of level-2 units, we summed the number of flagged statistics and sorted the table in descending order, thus, placing level-2 units most frequently identified as influential at the top of the summary table. In addition, the MIXED_DX influence output is calculated at the level-2 unit (controlled by the 'EFFECT =' option in Influence command in PROC MIXED) and based on five iterations (controlled by the 'ITER =' option in the Influence command in PROC MIXED).

MACRO MIXED_DX

The MIXED_DX macro inputs consist of ODS table names generated from the MIXED procedure including (a) ModelInfo,, (b) Dimsensions, (c) SolutionF, (d) SolutionR, (e) Level1, and (f) Influence, as well as two user specified arguments, MIN_NJ and PR. MIN_NJ defines the smallest level-2 unit to be included in the within-unit box-and-whisker plots and normality assessments (allowing the exclusion of very small units from these analyses), whereas PR defines the threshold percentile rank of influence diagnostics that trigger the flagging of level-2 units. If the MIN_NJ subanalysis is not wanted, the user will need to override the default value of five by specifying some value for MIN_NJ such as MIN_NJ = 1.

The MIXED_DX macro produces numeric (i.e., output located in the output window) and visual (i.e., output located in the graph window) output for both level-1 and level-2 data.

Level-1 numeric output includes: (a) normality summary statistics table for the overall level-1 residual and for each level-2 unit, (b) normality summary statistics table for the overall level-1 residual and for each level-2 unit where $\boldsymbol{n}_j$ > min_nj, (c) PROC UNIVARIATE plot output for the variance of all level-1 residuals, (d) PROC UNIVARIATE plot output for the variance of level-1 residuals where $\boldsymbol{n}_j$ > min_nj, and (e) PROC GLM output providing Levene's Test for homogeneity of variance of level-1 residuals.

Level-1 visual output includes: (a) side-by-side box-and-whisker plots for the overall level-1 residual and for each level-2 unit, (b) side-by-side box-and-whisker plots for the overall level-1 residual and for each level-2 unit where $\boldsymbol{n}_j$ > min_nj, (c) histogram of the variance of all level-1 residuals, (d) histogram of the variance of all level-1 residuals where $\boldsymbol{n}_j$ > min_nj, and (e) plot of level-1 residuals by predicted values.

Level-2 numeric output includes: (a) PROC GLM output providing Levene's Test for homogeneity of variance of level-2 residuals for both the intercept and slope (when applicable) for each level-2 effect, (b) PROC UNIVARIATE plot output of level-2 residuals for each level-2 effect, (c) multivariate normality and outlier summary table, (d) Mahalanobis distance values for level-2 units, (e) PROC UNIVARIATE plot output of Mahalanobis distance values for level-2 units, and (f) table of ranked influence diagnostics.

Level-2 visual output includes: (a) plot of level-2 residuals by predicted values for each level-2 effect, (b) histogram of level-2 residuals for each level-2 effect, and (c) histogram of Mahalanobis distance values.

Below is important information about the types of models that are supported by macro MIXED_DX:

A. MIXED_DX has been tested for use in SAS 9.1.3 and 9.2
B. MIXED_DX utilizes the following SAS components: Base SAS, SAS/STAT, SAS/GRAPH, and SAS/IML
C. MIXED_DX works for 2-level linear multilevel models estimated in PROC MIXED
D. MIXED_DX works correctly for models that have converged and have positive definite G-matrices
E. MIXED_DX is designed to read model effect names up to 32 characters in length. If a user has an effect that is longer than 32 characters, the 'NAMELEN= ' option in PROC MIXED must be updated accordingly.
F. If a user's model contains any interactions, they must be created on the model statement (e.g., model y=x1 x2 w1 w2 x1*w1)
G. Users cannot use the class statement to create dummy variables in PROC MIXED. That is, any dummy variables used in the model statement need to be created prior to the MIXED procedure (e.g., in a data step).

H. Users need to specify their desired value for MIN_NJ, which defines the smallest level-2 unit to be included in the within-unit box-and-whisker plots and normality assessments. The default is set to 5. Even if this subanalysis is not wanted, the user must override the default value of 5 and specify some value for MIN_NJ such as min_nj = 1.

I. For nominal level-2 effects that include multiple dummy vectors, users need to specify the dummy vector variable names, as they are listed in the PROC MIXED model statement, separated by spaces, for the 'GroupVar =' argument to the macro; the default is 'none'. Only one set of dummy vectors can be entered for the 'GroupVar' argument; if the PROC MIXED model statement contains more than one level-2 multi-group variable that has been dummy coded, the macro needs to be run separately for each of these effects to generate the Levene's Test for homogeneity of variance for each specific multi-group level-2 effect. If the PROC MIXED model statement does not include any level-2 dummy coded variables, then the user needs to keep the default value of 'none'.

J. Users need to specify their desired value for PR, which defines the threshold percentile rank of influence diagnostics that trigger the flagging of level-2 units. The default is set to 90.

The most current version of the complete MIXED_DX SAS program file is available for download from http://www.ed.sc.edu/bell/ .

EXAMPLE OF MACRO MIXED_DX

Below is an example of PROC MIXED code used in conjunction with the macro MIXED_DX, followed by select examples of the numeric and visual output.

```
ods exclude influence SolutionR;
title;
proc mixed data = temp covtest noclprint NAMELEN=32;
class schoolid;
model mathach = SIZE FEMALE SES MEANSES SES*MEANSES SES*SIZE
FEMALE*SIZE/ solution outp=L1Resid influence(effect=schoolid iter=5);
random  intercept FEMALE SES / sub=schoolid solution type=un;
ods output SolutionR=L2Resid SolutionF=Fixed ModelInfo=ModelStuff
Dimensions=DatStuff Influence=influence; run;
%mixed_dx (ModelI = modelstuff, Dims = datstuff, solnF = Fixed, solnR =
L2Resid, Level1 = L1Resid, Influence = influence, min_nj = 5, pr = 90,
GroupVar = None); run;
```

MIXED_DX VISUAL OUTPUT

Figure 1 contains one 'window' of the level-1 residual box-and-whisker plot output for the overall level-1 residual (the first plot in the series) and for each level-2 unit. These plots can be used to evaluate the normality and heterogeneity of variance assumption for the overall level-1 residual, as well as for the level-1 residuals within each level-2 unit. Figure 2 contains similar output as Figure 1, however, instead of including box-and-whisker plots for all level-2 units, the output in Figure 2 displays the overall level-1 residual and the residuals for level-2 units in which nj ≥ 5 , where the sample size of five was specified through the MIN_NJ = 5 option in the call to the macro. As with Figure 1, the box-and-whisker plots included in Figure 2 allow an analyst to evaluate the tenability of the normality and heterogeneity of variance assumptions for level-1 residuals. However, instead of examining level-1 residuals within each level-2 unit, this part of the MIXED_DX output is limited to level-2 units with a user specified minimum number of observations. This approach allows users to focus on level-2 units with enough observations that normality and equal variances might be plausible (i.e., when level-2 units do not have many observations in them, normally distributed errors and equal variances with other, more dense level-2 units is highly improbable given the tenets of Central Limit Theorem).

The histogram displaying the distribution of the variance of level-1 residuals for all level-2 units shown in Figure 3 provides users with summary information regarding the magnitude of level-1 residual variances across level-2 units. Although this part of the output is not directly related to evaluating the heterogeneity of variance assumption for level-1 residuals, it offers supplemental information regarding the nature of the variability in level-1 residuals. Figure 4 displays the plot of level-1 residuals by predicted values for all observations included in the PROC MIXED analysis. Similarly, Figures 5 and 6 contain plots of level-2 residuals by predicted values for the intercept and the random effect for the level-1 variable 'female' (i.e., level-2 effect = female), respectively. Each scatterplot provides users a way to evaluate simultaneously the normality, linearity, and homoscedasticity of level-1 (Figure 4) and level-2 residuals (Figures 5 and 6).

Figure 7 contains a histogram displaying the distribution of Mahalanobis distance values for each level-2 unit which allows researchers to easily examine the range of values to determine if potentially problematic outliers are evident.

Moreover, when evaluating data for possible outliers or violations of statistical assumptions, nicely summarized distributions such as the histogram in Figure 7 are particularly helpful and efficient when there are oodles of level-2 units.

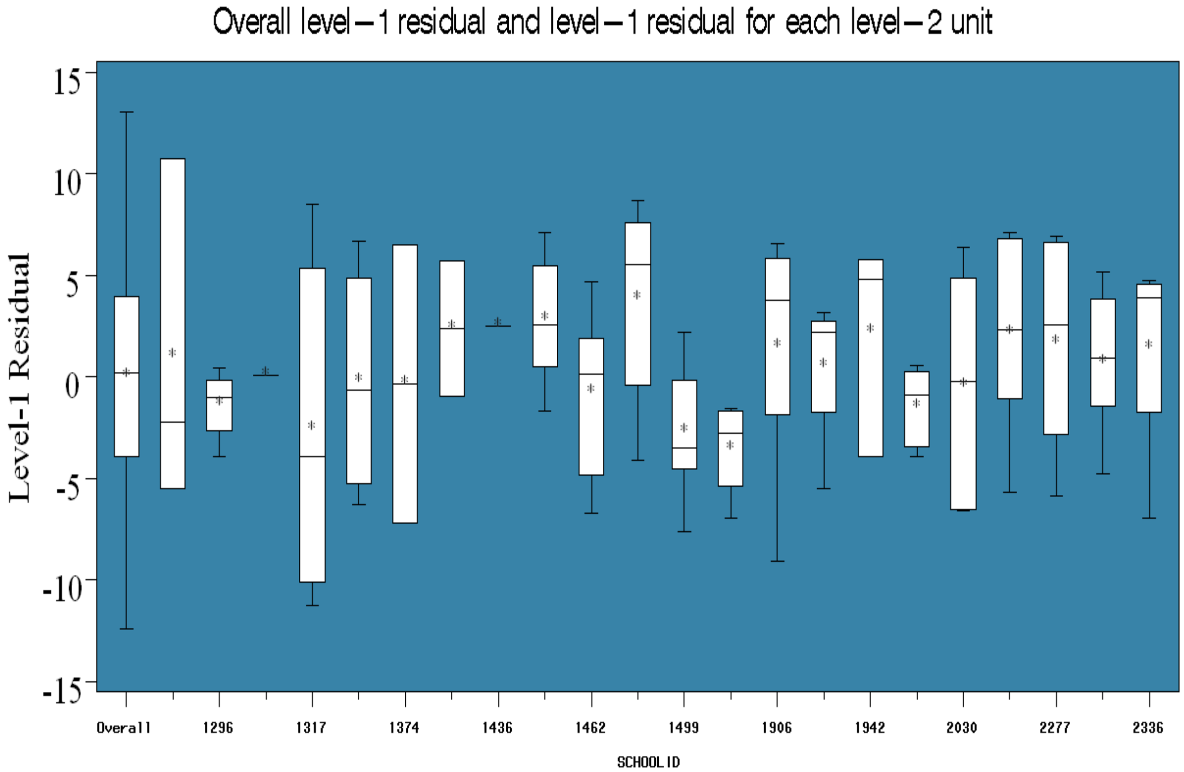

Figure 1. Distribution of level-1 residuals: Overall residual and for each level-2 unit

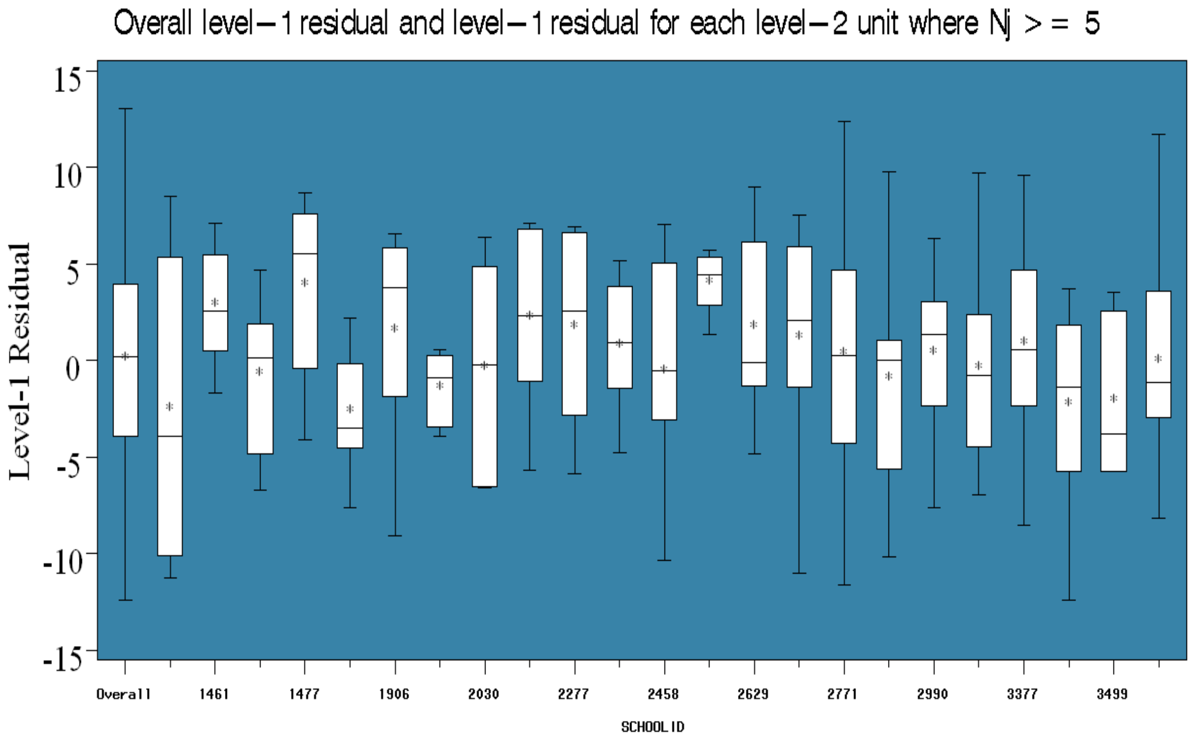

Figure 2. Distribution of level-1 residuals: Overall residual and for each level-2 unit where $n_j \geq 5$

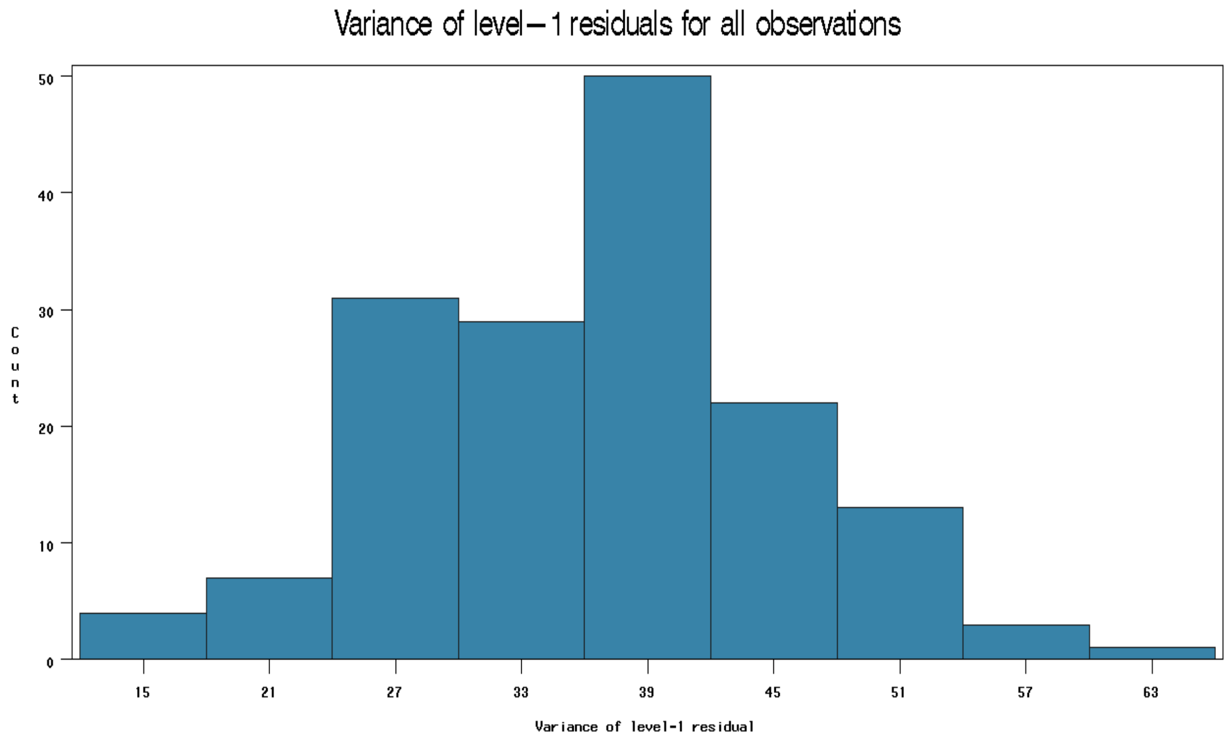

Figure 3. Distribution of the variance of level-1 residuals for all observations

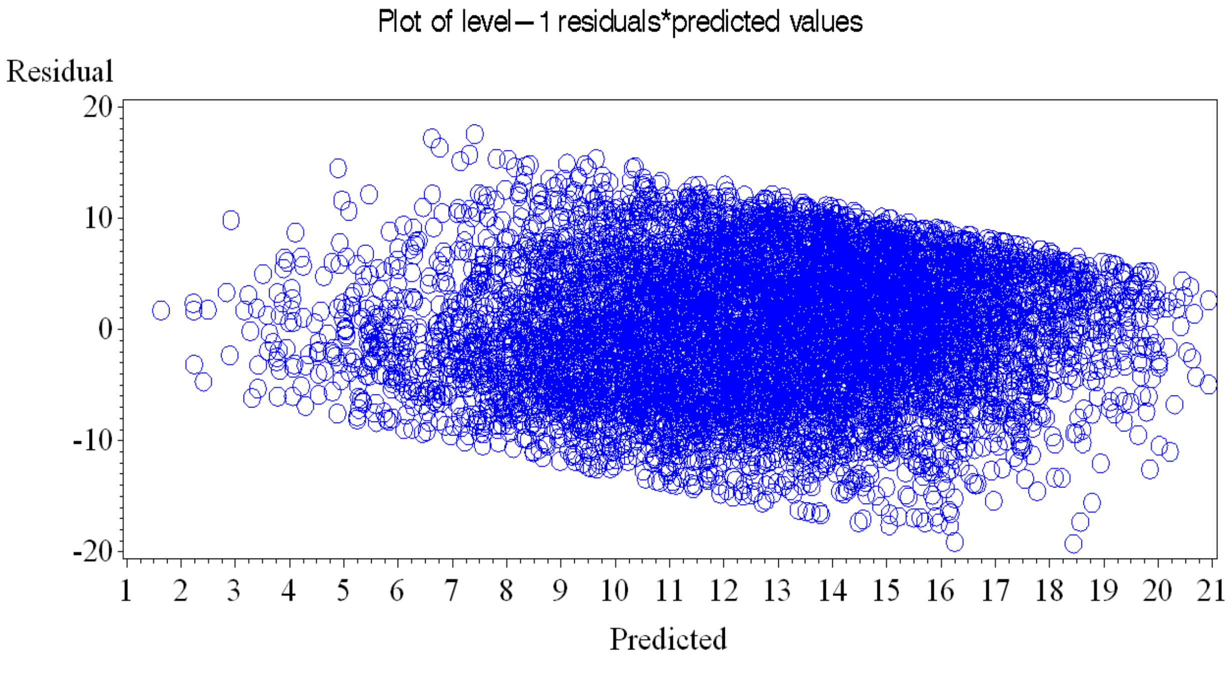

Figure 4. Level-1 residual*predicted value for all observations

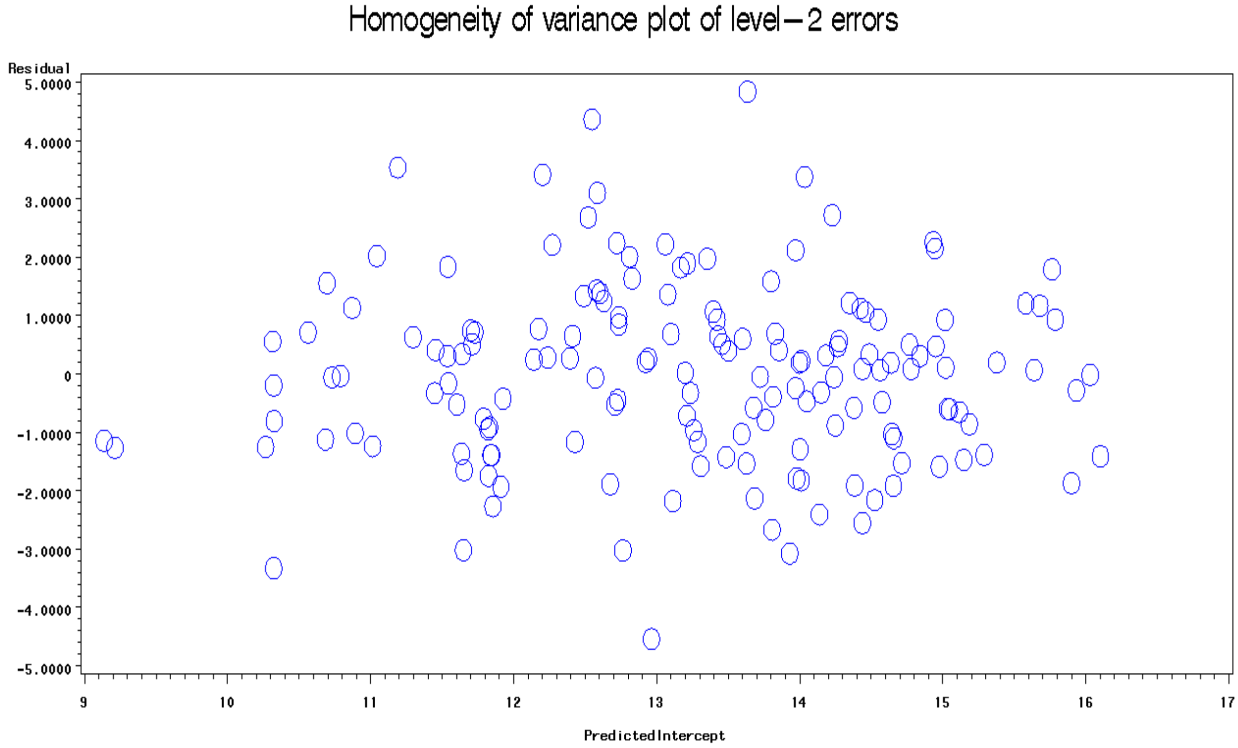

Figure 5. Level-2 residual*predicted intercept value for all level-2 units

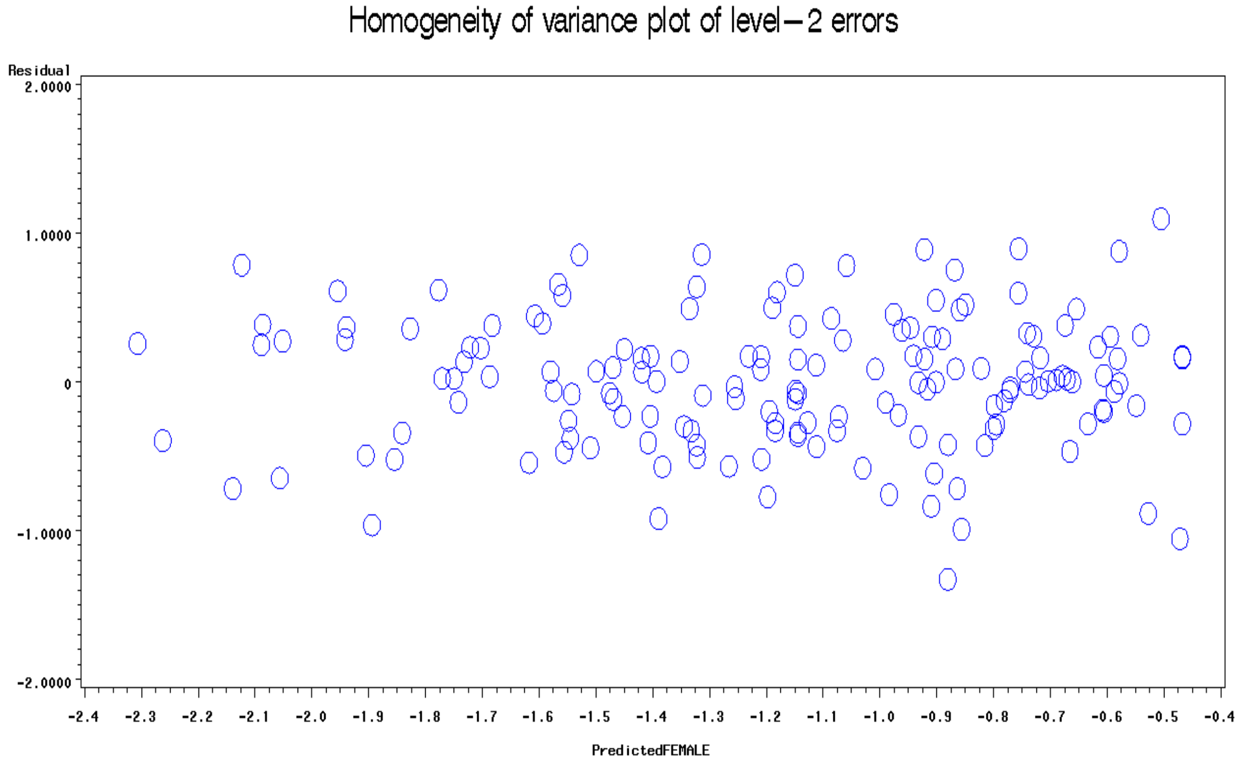

Figure 6. Level-2 residual*predicted random effect for level-1 variable ‘female’ for all level-2 units

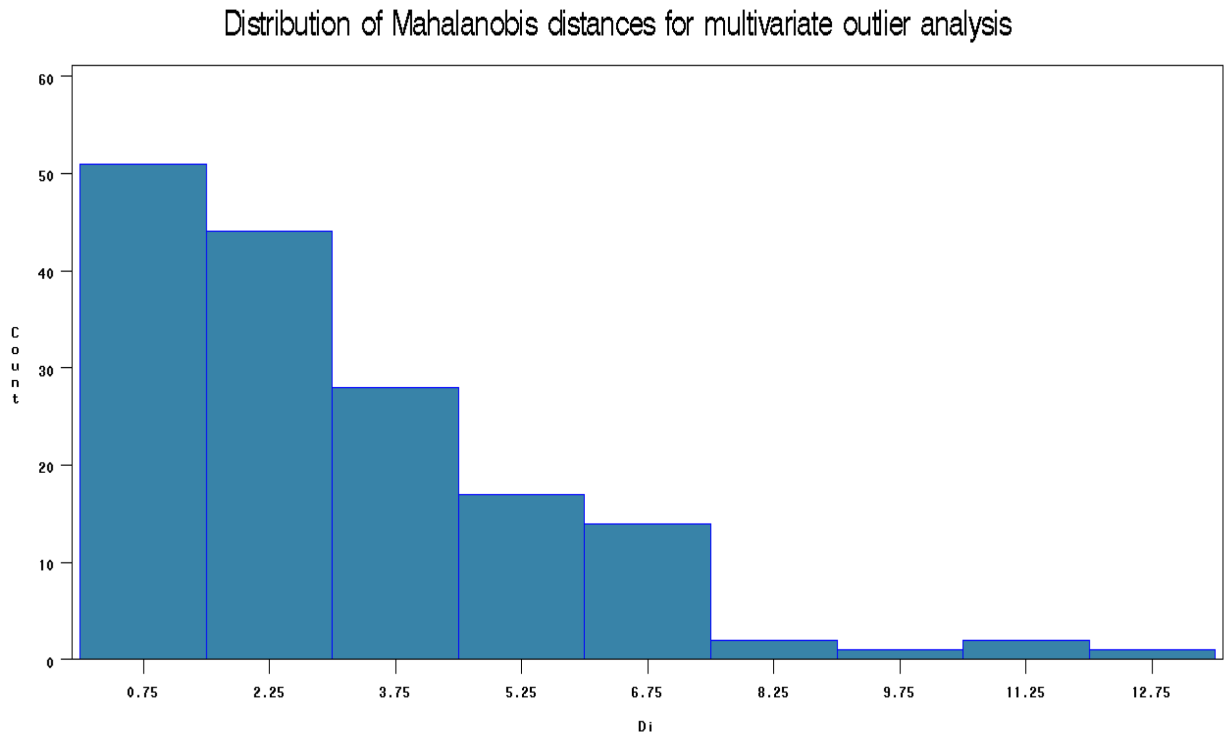

Figure 7. Distribution of Mahalanobis distance values for each level-2 unit

MIXED_DX NUMERICAL OUTPUT

Figures 8, 9, 10, 11, and 12 contain different elements of the numerical output generated by MIXED_DX. Specifically, Figure 8 is one 'window' of the normality summary table for level-1 residuals, Figure 9 contains the PROC GLM output for Levene's test of homogeneity of variance of level-1 residuals, Figure 10 includes PROC GLM output for the Levene's test of homogeneity of variance of level-2 residuals of the variable 'Size', Figure 11 includes the level-2 multivariate normality and outlier assessment table, and Figure 12 includes one 'window' of the ranked influence diagnostic summary table for level-2 units. The statistical output provided in the normality summary table for level-1 residuals (Figure 8) can be used to supplement the visual evaluation of the normality and homogeneity of variance assumption of level-1 residuals. Similarly, the Levene's test of homogeneity of variance output provided in Figures 9 and 10 can also be used to supplement any visual assessment of equal variances of (a) level-1 residuals across level-2 units and (b) level-2 intercept and slope residuals, respectively. The statistical output in Figure 11 allows users a method of evaluating the multivariate normality of level-2 residuals. Also, results from the inferential examination of the largest Mahalanobis distance can be used to supplement the visual evaluation of potentially problematic outliers. By providing a nice summary of the influential nature of each level-2 unit, the ranked influence diagnostics output in Figure 12 provides users a way to easily examine which, if any, level-2 units might be influencing the magnitude or precision of fixed and/or random effects in their model.

CONCLUSIONS

The macro MIXED_DX was created to facilitate the evaluation of statistical assumptions of linear two-level multilevel models, including the examination of residual normality, linearity, and homogeneity of variance, as well as influential outliers. By utilizing data provided in various ODS tables in PROC MIXED, MIXED_DX makes it much easier for the applied data analyst to evaluate assumptions for both level-1 and level-2 residuals, through both numeric and visual assessments such as statistical summary tables, box-and-whisker plots, histograms, and plots of residuals by predicted values. Furthermore, MIXED_DX is easily modifiable. For example, users can add to the macro to invoke procedures that are not currently included, such as examining the distribution of level-2 sample sizes. Conversely, users can also easily comment out any unwanted elements of the macro. Thus, given the ease in calling the macro when estimating two-level linear models via PROC MIXED, we hope that more researchers will not only evaluate the distributional assumptions and outliers, but also report said evaluations in their published research.

Normality summary statistics for the overall residual and each level-2 unit

Level-2 Unit	Skewness	Kurtosis	Shapiro-Wilk	p-value	Variance	Std Dev	N
Overall	-0.1464	-0.5220	0.0321	0.0100	35.6292	5.9690	7185
1224	0.2481	-1.1449	0.9375	0.0143	54.4537	7.3793	47
1288	-0.4288	-0.6287	0.9475	0.2206	46.0958	6.7894	25
1296	0.6568	1.3472	0.9562	0.0711	28.5501	5.3432	48
1308	-0.3487	-0.5644	0.9483	0.3423	38.2406	6.1839	20
1317	0.1423	-0.8975	0.9604	0.1052	29.4795	5.4295	48
1358	-0.4556	0.2699	0.9683	0.4938	25.4989	5.0496	30
1374	-0.3089	-0.7325	0.9599	0.3471	61.3113	7.8302	28
1433	-0.4066	-0.7341	0.9572	0.1883	13.8575	3.7226	35
1436	-0.7585	0.1641	0.9476	0.0450	17.8752	4.2279	44
1461	-1.4369	2.4512	0.8717	0.0011	34.2199	5.8498	33
1462	0.5766	-0.4901	0.9467	0.0139	41.0545	6.4074	57
1477	-0.4332	-0.7634	0.9522	0.0171	50.5179	7.1076	62
1499	0.6850	0.0731	0.9523	0.0337	32.1495	5.6701	53
1637	0.7102	-0.0295	0.9374	0.1050	45.9021	6.7751	27
1906	-0.5982	-0.1637	0.9573	0.0561	40.8894	6.3945	53
1909	-0.2498	-1.3750	0.9193	0.0333	34.4495	5.8694	28
1942	-0.9756	0.7643	0.8972	0.0084	30.7441	5.5447	29
1946	0.0692	-1.1729	0.9503	0.0843	39.2686	6.2665	39
2030	-0.0596	-0.7441	0.9765	0.4578	38.4415	6.2001	47
2208	-0.2361	-1.1965	0.9445	0.0087	33.6639	5.8021	60
2277	0.5186	-0.3404	0.9664	0.0916	33.4067	5.7799	61
2305	-0.5484	-0.0718	0.9647	0.0541	26.2418	5.1227	67
2336	-0.5344	-0.2465	0.9653	0.1742	32.5628	5.7064	47
2458	0.1452	-1.1178	0.9484	0.0167	30.7462	5.5449	57
2467	0.1245	-0.5803	0.9855	0.7739	44.0580	6.6376	52
2526	-0.5742	-0.1631	0.9460	0.0130	23.5049	4.8482	57
2626	-0.0066	-0.8930	0.9711	0.4212	30.8799	5.5570	38
2629	-0.4906	0.0390	0.9707	0.1803	27.7982	5.2724	57
2639	0.5922	0.0269	0.9667	0.2544	36.6968	6.0578	42
2651	-0.0105	-0.5696	0.9825	0.8043	39.6750	6.2988	38
2655	-0.1188	-0.9960	0.9674	0.1643	31.3419	5.5984	52
2658	-0.1122	-0.4767	0.9887	0.9346	29.0280	5.3878	45
2755	-0.4818	-0.7205	0.9536	0.0602	37.3001	6.1074	47
2768	0.2829	-0.2105	0.9819	0.9205	42.1063	6.4889	25
2771	0.1278	-0.7310	0.9700	0.1845	43.4792	6.5939	55
2818	-0.1103	-1.2276	0.9533	0.0849	27.7784	5.2705	42
2917	0.1837	-0.7458	0.9711	0.3458	49.9826	7.0698	43
2990	-0.8184	0.3007	0.9410	0.0177	18.3683	4.2858	48
2995	0.1756	-1.0986	0.9563	0.0825	51.9669	7.2088	46
3013	-0.3269	-0.9800	0.9541	0.0406	44.7080	6.6864	53
3020	-0.6004	0.2804	0.9745	0.2490	39.1896	6.2602	59
3039	-0.5377	0.4199	0.9710	0.7551	24.8429	4.9843	21
3088	0.0755	-0.8858	0.9691	0.3519	35.4447	5.9535	39
3152	-0.6367	-0.0941	0.9455	0.0188	46.7917	6.8404	52
3332	-0.0922	-0.8857	0.9646	0.2673	41.5297	6.4444	38
3351	0.1423	-0.8422	0.9773	0.6059	44.7192	6.6872	39
3377	0.2254	-0.6960	0.9772	0.5098	52.4213	7.2403	45
3427	-0.5809	-0.6186	0.9375	0.0118	14.0454	3.7477	49

Figure 8. Partial output from MIXED_DX normality summary table for level-1 residuals: Overall and for each level-2 unit

Levenes test of homogeneity of variance of level-1 residuals

The GLM Procedure

Dependent Variable: Absolute_resid (Absolute value of level-1 residual)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	159	3269.56971	20.56333	1.85	<.0001
Error	7025	77957.87122	11.09721		
Corrected Total	7184	81227.44092			

R-Square	Coeff Var	Root MSE	Absolute_resid Mean
0.040252	67.55124	3.331247	4.931437

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SCHOOLID	159	3269.569707	20.563331	1.85	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SCHOOLID	159	3269.569707	20.563331	1.85	<.0001

Figure 9. MIXED_DX output for Levene's homogeneity of variance test of level-1 residuals

This is Levenes Test for Discretized Level-2 Predictor SIZE

The GLM Procedure

Dependent Variable: Absolute_resid Absolute Value 'Intercept'

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7.4023303	1.8505826	2.08	0.0867
Error	155	138.2284266	0.8917963		
Corrected Total	159	145.6307569			

R-Square	Coeff Var	Root MSE	Absolute_resid Mean
0.050829	79.02124	0.944350	1.195058

Source	DF	Type I SS	Mean Square	F Value	Pr > F
D_Predict	4	7.40233027	1.85058257	2.08	0.0867

Source	DF	Type III SS	Mean Square	F Value	Pr > F
D_Predict	4	7.40233027	1.85058257	2.08	0.0867

Figure 10. MIXED_DX output for Levene's homogeneity of variance test of level-2 residuals of the variable 'size'

```
Multivariate normality and outlier assessment
---------------------------------------------
Multivariate Skewness:
         b1p =                 0.45811311
         Chi-Square =          12.5624477
         df =                          10
         p-value =             0.24917459

Multivariate Kurtosis:
         b2p =                 14.7863678
         Z (Upper) =           -0.2466812
         p-value (Upper) =     0.59742254
         Z (Lower) =           -0.6775666
         p-value (Lower) =     0.24902327

Largest Mahalanobis Distance:
         Level 2 Unit =             3427
         D-Squared =            13.135085
         F =                   4.71470232
         df numerator =                 3
         df denominator =             156
         p-value               0.00353276
```

Figure 11. MIXED_DX output for level-2 multivariate normality and outlier assessment

Ranked influence diagnostics

Obs	SCHOOLID	flags	Nobs	Iter	PRESS	MDFFITS	CovRatio	CovTrace	CookDCP	MDFFITSCP	CovRatioCP	CovTraceCP	RMSE	RLD
1	3427	8	49	2	2406.76	0.01810	0.8530	0.1512	0.53129	0.60558	0.7554	0.2546	6.05025	0.7456
2	8367	8	14	2	1244.06	0.02893	0.8735	0.1305	0.25961	0.30125	0.7396	0.2786	6.03687	0.4877
3	2277	7	61	2	2273.07	0.04285	0.9057	0.0948	0.25883	0.28665	0.7984	0.2094	6.04184	0.6579
4	2305	7	67	2	2192.31	0.02412	0.9872	0.0110	0.12724	0.12935	0.9405	0.0570	6.04602	0.3455
5	6074	7	56	2	2990.11	0.02145	0.9742	0.0239	0.12442	0.13011	0.9222	0.0756	6.03473	0.3118
6	2655	6	52	2	2623.47	0.04934	1.0959	0.0982	0.13220	0.14074	0.9957	0.0044	6.03880	0.5516
7	6990	6	53	2	2177.66	0.02582	0.9600	0.0364	0.18515	0.20090	0.9699	0.0234	6.04228	0.3824
8	3533	5	48	2	2500.44	0.01383	0.9214	0.0779	0.29336	0.32961	0.8320	0.1706	6.03095	0.4105
9	3716	5	41	2	2114.35	0.03330	1.0337	0.0361	0.12585	0.13249	0.9420	0.0555	6.03160	0.4154
10	8628	5	61	2	3382.65	0.01206	0.9408	0.0570	0.21589	0.22709	0.9976	0.0072	6.03865	0.3254
11	2526	4	57	2	1873.14	0.01222	0.9795	0.0192	0.09784	0.10099	0.9773	0.0187	6.04531	0.1929
12	2639	4	42	2	1574.98	0.03415	1.2044	0.1955	0.03084	0.03170	1.0291	0.0310	6.03686	0.3174
13	7345	4	56	2	2558.10	0.00704	0.9630	0.0353	0.15047	0.15451	1.0219	0.0252	6.03364	0.2177
14	7688	4	54	2	2021.59	0.00813	0.9567	0.0428	0.12721	0.13235	0.9245	0.0734	6.04651	0.2013
15	8193	4	43	2	1573.01	0.01828	0.9898	0.0095	0.09397	0.09611	0.9133	0.0874	6.04648	0.2618
16	8627	4	53	2	2543.96	0.01852	1.1416	0.1364	0.03256	0.03279	1.0667	0.0658	6.03346	0.1939

Figure 12. Partial output from MIXED_DX ranked influence diagnostics summary table for each level-2 unit

REFERENCES

Andrews, D.F. & Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society, Series B, 40,* 85-93.

Belsley, D.A., Kuh, E., & Welsch, R.E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* New York: Wiley.

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics, 19,* 15-18.

Cook, R.D. & Weisberg, S. (1982). *Residuals and influence in regression.* New York: Chapman and Hall.

Ferron, J.M., Hogarty, K.Y., Dedrick, R.F., Hess, M.R., Niles, J.D., & Kromrey, J.D. (2008). Reporting results from multilevel analyses. In A.A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 391-426). Charlotte, NC: Information Age Publishing, Inc.

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

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R., D. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute Inc.

Longford, N. (2001). Simulation-based diagnostics in random-coefficient models. *Journal of the Royal Statistical Society Series A-Statistics in Society, 164,* 259-273.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Schabenberger, O. (2004). *Mixed model influence diagnostics*. *Proceedings of the 29th Annual SAS Users Group International Conference*.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bethany A. Bell
University of South Carolina
College of Education, Wardlaw 133
Columbia, SC 29208
Work Phone: 803-777-2387
E-mail:babell@sc.edu
Web: http://www.ed.sc.edu/bell/