Robust Standard Errors

The maximum likelihood based estimation used with multilevel regression for continuous variables leads to particular concern about the normality assumption for the fixed effects tests, because nonnormal data can lead to incorrect standard error estimates, and, thus, significance tests. The nature of the impact on standard errors depends on the shape of the distribution. Fortunately, adjustments to the standard error estimation have been developed that work well in many situations (Burton, Gurrin, & Sly, 1998; Eiker, 1963; Huber, 1967; Liang & Zeger, 1986; White, 1982). Robust estimates may perform best when there are 100 level-2 units (groups) or more (Cheong, Fotiu, & Raudenbush, 2001; Hox & Maas, 2001; Krauermann & Carroll, 2001). The robust standard errors are known as Huber-White or Huber-White-Eiker or "sandwich" estimation. These adjustments also appear to be helpful for heteroscedasticity (Beck & Katz, 1997). There may be a slight cost in power with these adjustments (robust estimates will tend to be slightly larger than standard asymptotic estimates; Hox, Moerbeek, & van de Schoot, 2018), but, with sufficient number of groups, the minimal power loss is probably less of a concern than nonnormality. Other robust estimation approaches exist. One, called bias reduced linearization (BRL) or CR2 (Bell & McCaffrey, 2002)[1] seems to work well with a small number of groups and corrects for heteroscedasticity (Huang & Li, 2022; Huang, Wiederman, & Zhang, 2023).

Robust standard errors are available in several statistical packages, including HLM (included in the output by default under “Robust”), SAS PROC MIXED (“Empirical”), and Stata (“Huber-White Sandwich”). Version 19 and higher of SPSS includes robust estimation only with the GENLINMIXED command. The `MLMusingsR` package in R can be used in conjunction with the `lme4` package.

I use the HSB model with the cross-level interaction between SES and sector to illustrate. The residuals for math achievement do not seem to be terribly nonnormal. And given the large number of groups and total sample size, we should not expect to see large differences in the standard errors between the usual (model-based) standard errors and the robust errors. There also seem to be some minor discrepancies in the standard error values across the three packages I illustrate below.

SPSS

Beginning with Version 19, SPSS provides robust standard error estimates in the GENLINMIXED procedure (but not with MIXED). The GENLINMIXED procedure is less user friendly. It is designed to be used with non-continuous outcomes[2] but can provide robust standard errors for a model with a continuous outcome. To obtain robust standard errors, I changed the default DF with DF_METHOD=SATTERTHWAITE and I requested robust (Huber-White) standard errors with COVB=ROBUST. For smaller number of groups (e.g., 50 or fewer), I recommend using the Kenward-Roger degrees of freedom, specifying DF_METHOD= KENWARD_ROGER. With the Kenward-Roger degrees of freedom, SPSS uses the regular (MODEL) based standard errors (and will indicate this in the output). Since these are the default standard errors, you can just remove the COVB statement when using Kenward-Roger for small sample size.

```
* GENLINMIXED requires that the ID variable be declared as nominal level (mixed does not appear to require this).

VARIABLE LEVEL schoolid (NOMINAL).

*REML model to get robust estimates.

GENLINMIXED
/DATA_STRUCTURE SUBJECTS=schoolid
/FIELDS TARGET= mathach
/TARGET_OPTIONS DISTRIBUTION=NORMAL LINK=IDENTITY
/BUILD_OPTIONS DF_METHOD=SATTERTHWAITE COVB=ROBUST
/FIXED  EFFECTS= cses sector cses*sector USE_INTERCEPT=TRUE
/RANDOM EFFECTS=cses USE_INTERCEPT=TRUE SUBJECTS=schoolid
COVARIANCE_TYPE=UNSTRUCTURED.
```

[1] The `clubsandwich` package in R can obtain CR2 standard errors, https://cran.r-project.org/web/packages/clubSandwich/index.html.
[2] More on this topic later.

The TARGET is the dependent variable. DISTRIBUTION=NORMAL LINK=IDENTITY is used to request the REML estimation for a continuous variable (note that ML is not currently available with GENLINMIXED).

The output by default contains some creative diagrams of the results, which was the only output by default in earlier versions, but the most recent versions of SPSS seem to also print the "table" output as well. The standard error adjustments are not noted anywhere in the output.

Fixed Coefficients [a]

Model Term	Coefficient	Std. Error	t	Sig.	95% Confidence Interval Lower	Upper
Intercept	11.411	.2927	38.984	<.001	10.833	11.989
cses	2.803	.1579	17.753	<.001	2.491	3.115
sector	2.800	.4359	6.422	<.001	1.938	3.661
cses*sector	-1.341	.2305	-5.819	<.001	-1.797	-.886

Probability distribution: Normal
Link function: Identity
a. Target: mathach

R

The model is tested as usual with `lme4` and then the `MLMusingsR` can be used with the model results from `lme4`. The estimation with `MLMusingsR` make take a minute or two.

```
> library(lme4)
> model1 <- lmer(mathach ~ cses + sector + cses*sector + (cses|schoolid), data = mydata, REML = TRUE)
> summary(model1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: mathach ~ cses + sector + cses * sector + (cses | schoolid)
   Data: mydata

REML criterion at convergence: 46638.6

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.06490 -0.73237  0.01565  0.75370  2.94195

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 schoolid (Intercept)  6.7504  2.5982
          cses         0.2657  0.5154   0.78
 Residual             36.7056  6.0585
Number of obs: 7185, groups:  schoolid, 160

Fixed effects:
            Estimate Std. Error       df t value             Pr(>|t|)
(Intercept)  11.4106     0.2930 158.4267  38.944 < 0.0000000000000002 ***
cses          2.8028     0.1550 141.6607  18.087 < 0.0000000000000002 ***
sector        2.7995     0.4395 153.7010   6.369        0.00000000209 ***
cses:sector  -1.3411     0.2338 151.5366  -5.737        0.00000005077 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cses   sector
cses         0.262
sector      -0.667 -0.175
cses:sector -0.174 -0.663  0.264
> library(MLMusingR)
> robust_mixed(model1)

Standard error type = CR2
Degrees of freedom = Satterthwaite

            Estimate  mb.se robust.se t.stat    df               Pr(>t)
(Intercept)   11.411  0.293     0.294 38.770  88.9 <0.0000000000000002 ***
cses           2.803  0.155     0.159 17.642  78.6 <0.0000000000000002 ***
sector         2.800  0.440     0.439  6.381 149.2 <0.0000000000000002 ***
cses:sector   -1.341  0.234     0.232 -5.777 137.5 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

HLM

Below, the HLM output for the cross-level interaction model allows for a comparison of the usual standard errors and the robust standard errors. In this case, the standard errors are highly comparable, but in other cases there may be more substantial differences in standard errors and significance levels. Conclusions may be different, and if there is a sufficient number of groups, I would trust the robust estimates more. If the number of groups is small, I would be more cautious about using the robust estimates. I include the standard estimates here for the sake of comparison. Notice that the estimates are the same, but the standard errors differ slightly. The degree of difference will depend on the degree of departure from normality of the dependent variable.

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	11.393836	0.292784	38.915	158	<0.001
SECTOR, γ_{01}	2.807465	0.439216	6.392	158	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.802449	0.156523	17.904	158	<0.001
SECTOR, γ_{11}	-1.340634	0.236028	-5.680	158	<0.001

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	11.393836	0.292348	38.974	158	<0.001
SECTOR, γ_{01}	2.807465	0.435634	6.445	158	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.802449	0.157937	17.744	158	<0.001
SECTOR, γ_{11}	-1.340634	0.230324	-5.821	158	<0.001

References

Beck, N., & Katz, J.N. (1997). The analysis of binary time-series-cross-section data and/or the democratic peace. Paper, *Annual Meeting of the Political Methodology Group*, Columbus, Ohio, July, 1997.
Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology, 28*, 169–182.
Burton, P., Gurrin, L., & Sly, P. (1998). Extending the simple regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modeling. *Statistics in Medicine, 17*, 1261–1291.
Cheong, Y.F., Fotiu, R.P., & Raudenbush, S.W. (2001). Efficiency and robustness of alternative estimators for two- and three-level models: The case of NAEP. *Journal of Educational and Behavioral Statistics, 26*, 411–429.
Eicker, F. (1963). Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions. *The Annals of Mathematical Statistics. 34* (2): 447–456
Hox, J., Moerbeek, M., & Van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications*. Routledge.
Hox, J.J., & Maas, C.J.M. (2001). The accuracy of multilevel structural equation modeling with pseudo-balanced groups and small samples. *Structural Equation Modeling, 8*, 157–174.
Huang, F. L., & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods, 54*, 1181–1199 1-19.
Huang, F. L., Wiedermann, W., & Zhang, B. (2022). Accounting for heteroskedasticity resulting from between-group differences in multilevel models. *Multivariate Behavioral Research, 58*(3), 637-657.
Huber, P. J. (1967, June). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (*Vol. 1*, No. 1, pp. 221-233).
Kauermann, G., & Carroll, R.J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association, 96*, 1387–1396.
Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13-22.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50*, 1–25.