Generalized Mixed Models for % Correct using SAS GLIMMIX and STATA MLEOGIT

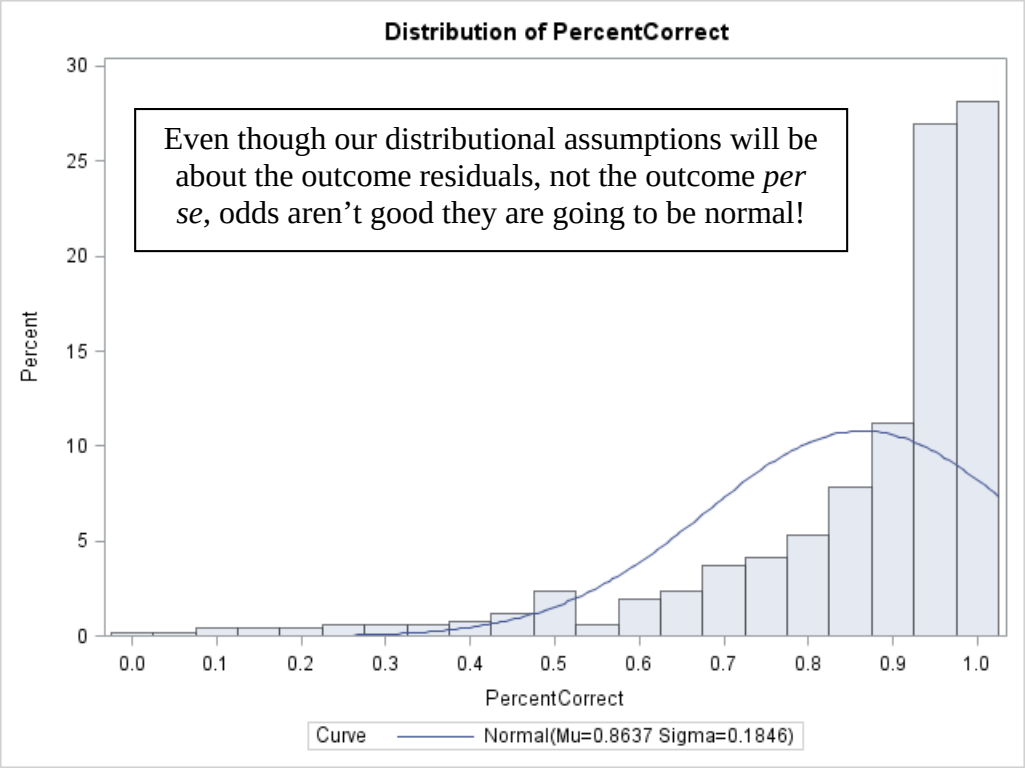
The data for this example are based on the publication below, which examined annual growth in a test of grammatical understanding from Kindergarten through 4th grade in children with non-specific language impairment (NLI) or specific language impairment (SLI). Given that percent correct is bounded between 0 and 1, we will use a logit link and a binomial response distribution for the level-1 residuals. However, given that the binomial is a discrete distribution, we will model the number of correct responses out of the number of trials directly instead.

Rice, M. L., Tomblin, J. B., **Hoffman, L.**, Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech-Language-Hearing Research, 47*(4), 816-834.
Available at: http://www.lesahoffman.com/Research/004%202004%20Rice%20et%20al.%20JSLRH.pdf .

SAS Data Manipulation:

```
* Reading in data, creating events and trials variables from % correct;
DATA work.growthdata; SET growth.grw2grps;
       Ntrials=100;
       Ncorrect=ROUND(PercentCorrect*Ntrials,1); RUN;

TITLE "Distribution of Percent Correct";
PROC UNIVARIATE NOPRINT DATA=growthdata; VAR PercentCorrect;
       HISTOGRAM PercentCorrect / MIDPOINTS= 0 TO 1 BY .05 NORMAL(MU=EST SIGMA=EST); RUN; QUIT;
PROC MEANS DATA=growthdata; VAR PercentCorrect; RUN; TITLE;
```



STATA Data Manipulation:

```
* Import data
use "$filesave\growthdata.dta", clear
drop group
gen group=class
label define fgroup 2 "NLI" 3 "SLI"
label values group fgroup
display as result "Distribution of Percent Correct"
summarize percentcorrect
hist percentcorrect
```

Single-Level Empty Means Model for % correct using DV = Events/Trials Binomial Model

```
display as result "Empty Means Single-Level Model"
melogit ncorrect , binomial(ntrials),
estat ic, n(104),
nlcom 1/(1+exp(-1*(_b[_cons]))) // intercept in probability
estimates store Fit1            // save fit stats

TITLE "Empty Means Single-Level Model";
PROC GLIMMIX DATA=growthdata NOCLPRINT NOITPRINT METHOD=QUAD(QPOINTS=5) GRADIENT;
       CLASS ID group wave;
       MODEL Ncorrect/Ntrials = / SOLUTION LINK=LOGIT DIST=BINOMIAL;
       ESTIMATE "Intercept" intercept 1 / ILINK; * ILINK un-logits prediction;
RUN;
```

Note: I am using 5 quadrature points to match STATA's default. In practice SAS will choose how many are needed if you don't specify (as 5 is likely to be too few).

Fit Statistics	
-2 Log Likelihood	**14444.36**
AIC (smaller is better)	14446.36
AICC (smaller is better)	14446.36
BIC (smaller is better)	14450.59
CAIC (smaller is better)	14451.59
HQIC (smaller is better)	14448.01
Pearson Chi-Square	14699.20
Pearson Chi-Square / DF	**28.94**

To go from logits to predicted % correct:

$$\text{Prob}(y = 1) = \frac{\exp(1.8475)}{1 + \exp(1.8475)} = .8638$$

The sample average probability of getting each item correct is .8638. But this model has over-dispersion (too much variance, likely because we haven't incorporated the correlation among occasions from the same person).

Estimates							
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Mean	Standard Error Mean
Intercept	1.8475	0.01294	507	142.81	<.0001	**0.8638**	0.001522

Empty Means, Random Intercept Two-Level Model for % correct

```
display as result "Empty Means Random Intercept Two-Level Model"
melogit ncorrect , || id: , binomial(ntrials),
estat ic, n(104),
nlcom 1/(1+exp(-1*(_b[_cons]))) // intercept in probability
estimates store Fit2            // save fit stats
lrtest Fit2 Fit1                // LRT for random intercept variance

TITLE "Empty Means Random Intercept Two-Level Model";
PROC GLIMMIX DATA=work.growthdata NOCLPRINT NOITPRINT METHOD=QUAD(QPOINTS=5) GRADIENT;
       CLASS ID group wave;
       MODEL Ncorrect/Ntrials = / SOLUTION LINK=LOGIT DIST=BINOMIAL;
       RANDOM INTERCEPT / SUBJECT=ID TYPE=UN;
       ESTIMATE "Intercept" intercept 1 / ILINK; * ILINK un-logits prediction;
       COVTEST "Need Random Intercept?" 0;       * Test if random intercept is needed;
RUN;
```

COVTEST is a score test to evaluate the change in fit if parameters labeled as 0 were removed from the model.

Fit Statistics	
-2 Log Likelihood	**9304.03**
AIC (smaller is better)	9308.03
AICC (smaller is better)	9308.06
BIC (smaller is better)	9313.32
CAIC (smaller is better)	9315.32
HQIC (smaller is better)	9310.18

Fit Statistics for Conditional Distribution	
-2 log L(Ncorrect | r. effects)	8819.39
Pearson Chi-Square	7083.43
Pearson Chi-Square / DF	**13.94**

Covariance Parameter Estimates				
Cov Parm	Subject	Estimate	Standard Error	Gradient
UN(1,1)	ID	0.9647	0.1394	-0.001

The fixed intercept is not the same as in the previous single-level model because it is now conditional on the random intercept (expected proportions for someone with person mean = 0).

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Mean	Standard Error Mean
Intercept	2.1572	0.09787	103	22.04	<.0001	**0.8963**	0.009093

Tests of Covariance Parameters
Based on the Likelihood

Label	DF	-2 Log Like	ChiSq	Pr > ChiSq	Note
Need Random Intercept?	1	14444	5140.32	<.0001	MI

MI: P-value based on a mixture of chi-squares.

Random intercept 95% confidence interval in logits = 2.1572 ± 1.96*SQRT(0.9647) = 0.232 to 4.0823, which translates to predicted individual mean probabilities of getting an item correct of .558 to .983.

Saturated Means for Wave*Group, Random Intercept (Two-Level) Model for % Correct

```
display as result "Saturated Means by Wave and Group Random Intercept Two-Level Model"
melogit ncorrect i.wave##i.group, || id: , binomial(ntrials),
estat ic, n(104),
contrast i.wave#i.group                                // multivariate Wald tests
margins i.wave#i.group, predict(xb fixedonly)          // predicted logits
margins i.wave#i.group, predict(xb fixedonly) pwcompare(pveffects) // pairwise comparisons
margins i.wave#i.group, predict(   fixedonly)           // predicted probabilities

TITLE "Saturated Means by Wave*Group Random Intercept Two-Level Model";
PROC GLIMMIX DATA=work.growthdata NOCLPRINT NOITPRINT METHOD=QUAD(QPOINTS=5) GRADIENT;
      CLASS ID group wave;
      MODEL Ncorrect/Ntrials = wave|group / SOLUTION LINK=LOGIT DIST=BINOMIAL;
      RANDOM INTERCEPT / SUBJECT=ID TYPE=UN;
      LSMEANS wave*group / ILINK PLOT=MEANPLOT(SLICEBY=group JOIN);       * Plot in logts;
      LSMEANS wave*group / ILINK PLOT=MEANPLOT(SLICEBY=group JOIN ILINK); * Plot in prob;
      LSMEANS wave*group / DIFF=ALL;                                       * Pairwise diffs;
RUN;
```

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
wave	4	396	880.08	<.0001
group	1	396	6.49	0.0112
group*wave	4	396	27.44	<.0001

Fit Statistics

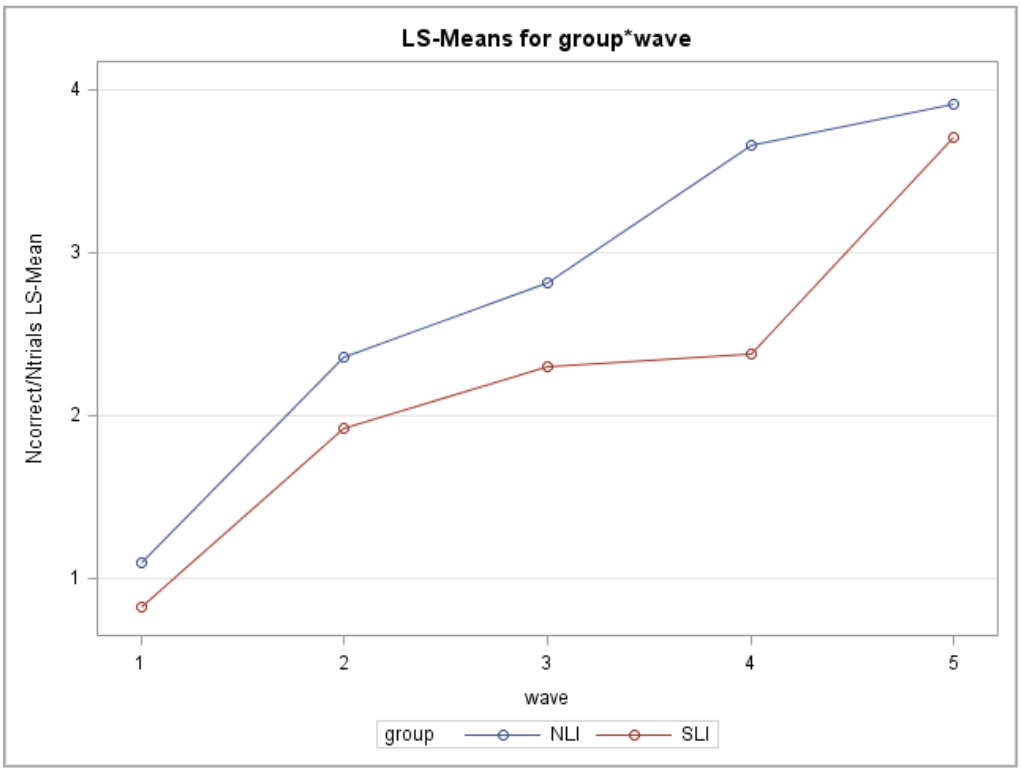
-2 Log Likelihood	5092.40
AIC (smaller is better)	5114.40
BIC (smaller is better)	5143.49

Fit Statistics for Conditional Distribution

-2 log L(Ncorrect | r. effects)	4601.63
Pearson Chi-Square	3291.09
Pearson Chi-Square / DF	**6.48**

group*wave Least Squares Means

group	wave	Estimate	Standard Error	DF	t Value	Pr > |t|	Mean	Standard Error Mean
NLI	1	1.0919	0.1457	396	7.50	<.0001	0.7487	0.02740
NLI	2	2.3612	0.1492	396	15.82	<.0001	0.9138	0.01175
NLI	3	2.8179	0.1521	396	18.53	<.0001	0.9436	0.008090
NLI	4	3.6637	0.1605	396	22.83	<.0001	0.9750	0.003910
NLI	5	3.9082	0.1643	396	23.79	<.0001	0.9803	0.003170
SLI	1	0.8226	0.1573	396	5.23	<.0001	0.6948	0.03335
SLI	2	1.9259	0.1596	396	12.06	<.0001	0.8728	0.01772
SLI	3	2.3018	0.1612	396	14.28	<.0001	0.9090	0.01333
SLI	4	2.3733	0.1619	396	14.66	<.0001	0.9148	0.01262
SLI	5	3.7087	0.1734	396	21.38	<.0001	0.9761	0.004050



Continued linear growth is only possible when modeled in logits (as plotted above).

Probability (as shown below as the "inverse linked" axis) is bounded at 1,
which translates into quadratic growth in the NLI group,
and perhaps cubic growth in the SLI group.