EdPsych/Psych/Stat 587
C.J. Anderson

R: Homework 5
Answer Key

1. *Report $-2LnLike$, number of parameters, AIC, BIC from lmer, BIC new from bic.hlm, $\hat{\tau}$s, $\hat{\sigma}^2$, R_1^2 and R_2^2 for Models (d) through (u) from computer labs 1 and 2. Report (a table of this information would be very nice).*

   See table below.

   (a) *Are there any differences in terms of "best" models when from BIC from lmer and BIC from bic.hlm?*

   With these data and model, BIC.new and BIC from lmer have the lowest values for the same random intercept model (i.e., Model j) and the same random intercept and slope model (Model s).

   (b) *Based on the information criteria, among the random intercept models, which is the "best" (or is there a unique one)?*

   There is not a unique one. AIC indicates the model (m) would be the best; whereas, both BICs indicate that model (j) is the best. As is typicaly, the BIC picks a simpler model. The next best models based the BICs is the same one picked by AIC. Also note that the next best model according to BICs (model (m)) is not that different from model (j).

   (c) *Based on the information criteria, among* ALL *models (random intercept and random slopes models), which is the "best" (or is there a unique one)?*

   In this case, there is a unique "best" model. Model (s) has the smallest value of for the AICs, BIC.lmer's and BIC.new's.

   (d) *What is the value of the harmonic mean used to compute R_2^2?*

   Harmonic mean:

$$\bar{x}_+ = 146/3.627 = 40.25$$

(e) *Which is the "best" model based on R^2 measures? Are the models with the better R^2's the same as the better/best models according to the information criteria?*

Models (n) and (o) have the largest $R_2^2 = .80$. Models(m,(n), (o), (s), (t) and (u) all have $R_1^2 = .45$, which is the largest value.

GRADING NOTE: If a student reports more than 2 decimal places, then there may be differences (i.e., no points off); however, beyond 2 decimal places is not particularly meaningful.

(f) *Interpret the values of R_1^2 and R_2^2 from model (s).*

- $R_1^2 = .45$: The proportional reduction in squared prediction error of Y_{ij}'s using group centered math, gender, grade, hours watching TV, etc. versus the model with no predictor variables is 45%.
- $R2_2^2 = .78$: The proportional reduction in squared prediction error of the group means $\bar{Y}_{+j}$ using predictor variables relative to the null model (i.e., no predictor variables) is 78%.

Model	$-2\times$ LnLike	#	AIC	BIC.lmer	BIC.new	$\hat{\tau}$s	$\hat{\sigma}^2$	R_1^2	R_2^2
From Lab 1/Homework 3									
Random Effects (Intercept) Models									
(d)	51489.0	3	51495.0	51515.6	51507.9	$\tau_0^2 = 20.95$	78.58	.00	.01
(e)	48482.4	4	48490.4	48517.8	48510.1	$\tau_0^2 = 6.02$	52.23	.42	.68
(f)	48482.4	4	48490.4	48517.8	48510.1	$\tau_0^2 = 6.02$	52.23	.42	.68
(g)	48646.8	4	48654.8	48682.2	48674.5	$\tau_0^2 = 21.68$	52.20	.26	.00
(h)	51267.2	4	51275.2	51302.7	51294.9	$\tau_0^2 = 3.05$	78.62	.18	.78
(i)	48426.4	5	48436.4	48470.7	48463.0	$\tau_0^2 = 3.72$	52.23	.44	.78
(j)	48363.7	7	48377.7	**48425.8**	**48418.0**	$\tau_0^2 = 3.76$	51.75	.44	.78
(k)	48359.0	10	48379.0	48447.6	48439.9	$\tau_0^2 = 3.61$	51.75	.44	.79
(l)	48363.5	8	48379.5	48434.4	48426.7	$\tau_0^2 = 3.76$	51.75	.44	.78
(m)	48347.2	9	**48365.2**	48427.0	48419.2	$\tau_0^2 = 3.67$	51.65	.45	.79
From Lab 2/Homework 4									
Random Intercept									
(n)	48338.6	12	48362.6	48445.0	48437.3	$\tau_0^2 = 3.38$	51.66	.45	.80
Random Intercept & Slopes									
(o)	48306.8	14	48334.8	48431.0	48415.5	$\tau_0^2 = 3.44$	50.90	.45	.80
						$\tau_1^2 = 0.73$			
						$\rho_{01} = .36$			
(p)	48314.1	11	48336.1	48411.6	48396.1	$\tau_0^2 = 3.75$	50.91	.44	.78
						$\tau_2^2 = 0.73$			
						$\rho_{10} = .41$			
(q)	48342.7	11	48364.7	48440.3	48424.7	$\tau_0^2 = 3.925$	51.43	.44	.77
						$\tau_4^2 = 0.16$			
						$\rho_{40} = -.37$			
(r)	48346.6	11	48368.6	48444.1	48428.6	$\tau_0^2 = 4.25$	50.93	.44	.76
						$\tau_5^2 = 0.02$			
						$\rho_{50} = -.58$			
(s)	48267.3	12	**48291.3**	**48373.8**	**48358.2**	$\tau_0^2 = 3.69$	50.91	.45	.78
						$\tau_1^2 = 0.21$			
						$\rho_{10} = 0.54$			
(t)	48266.7	13	48294.7	48390.8	48375.3	$\tau_0^2 = 3.68$	50.91	.45	.79
						$\tau_1^2 = 0.21$			
						$\rho_1^2 = 0.55$			
(u)	48265.1		48297.1	48407.0	48391.42	$\tau_0^2 = 3.67$	50.90	.45	.79
						$\tau_1^2 = 0.21$			
						$\rho_{01} = .54$			

2. *Consider base model from computer lab3. Do you need a random slope? (Be sure to report the statistical test that you use for this question).*

Test for random slope: $H_o : \tau_1^2 = \tau_{10} = 0$ versus H_a : no H_o.

Test statistic ($\mathcal{M}_o$ is the base model without random slope and $\mathcal{M}_a$ is the base model:

$$\begin{aligned}\lambda &= -2(\text{LnLike}\mathcal{M}_o) - \text{LnLike}\mathcal{M}_a)) \\ &= 48261 - 48255 \\ &= 5.91\end{aligned}$$

mixture of p-value from χ_1^2 and χ_2^2 equals .03; Reject H_o, the data support the conclusion that we need a random slope for group centered math.

3. *Compare and contrast the standard errors of parameters and results of significance tests for fixed effects when you use the model based versus the robust estimators of the standard errors. Which do you think is the best to use for testing fixed effects and why?*

Note: For this problem, just look at base model.

	Fixed		Model Based			Sandwiche/data based		
	Est.	satterthwaite	se.	t	p	se	t	p
(Intercept)	20.59	139.72	6.23	3.30	0.00	6.11	3.37	0.00
xSchoolCenterMath	30.49	137.03	3.31	9.20	0.00	3.13	9.73	0.00
gendergirl	-1.28	6979.87	0.17	-7.41	0.00	0.18	-6.96	0.00
grade4	0.87	6945.55	0.19	4.50	0.00	0.21	4.20	0.00
hoursTV	-0.09	7053.53	0.07	-1.23	0.22	0.08	-1.11	0.27
hourscomputergames	-0.26	7061.58	0.08	-3.27	0.00	0.08	-3.25	0.00
xSchoolMeanMath	4.05	141.57	0.18	22.90	0.00	0.18	22.18	0.00
typecommunity2	-3.25	114.95	2.08	-1.56	0.12	0.36	-8.89	0.00
typecommunity3	-4.25	114.06	2.05	-2.07	0.04	0.32	-13.30	0.00
typecommunity4	-4.33	113.95	2.05	-2.11	0.04	0.27	-16.26	0.00
xSchoolCenterMath:xSchoolMeanMath	-0.75	143.88	0.10	-7.77	0.00	0.09	-7.89	0.00
xSchoolCenterMath:typecommunity2	-0.38	86.98	0.97	-0.40	0.69	0.22	-1.73	0.09
xSchoolCenterMath:typecommunity3	-0.58	84.46	0.95	-0.61	0.55	0.17	-3.43	0.00
xSchoolCenterMath:typecommunity4	-0.88	83.92	0.94	-0.94	0.35	0.13	-7.04	0.00

- The robust standard errors for fixed effects tend to be of similar values for the model based ones, except for type of community and cross-level interactions involving. These are considerably different than the model based ones.
- The results of the significance tests for type of community change depending on which standard errors are used.

4. *Report what contrasts you tested, the results of them, and any action you took based on the results for*

I did the following using both empirical and model based standard errors; however, I report the model based ones below.

I fit models sequentially according to order of the models reported the table labeled "Model Refinements".

(a) *Type of community:* I did two contrasts

- H_1: $\gamma_{urban} = \gamma_{suburban}$, $F_{1,\sim 114} = .04$, p-value=.85. Could re-code these as one for main effects.
- H_2: $\gamma_{isolate} = \gamma_{rural}$ and the corresponding γ's for interactions with `(grpCmath)`. $t_1 15 = -1.561$, $p = .12$.
- H_3: average of γ_{urban} & $\gamma_{suburban} - \gamma_{rural}$. $F_{1,115} = 4.694$ and $p = .032$.

Given results of contrasts, I recoded location (i.e., typecommunity) such that: 1=isolated=rural, and 3=urban=suburban.

The likelihood ratio test of whether this re-coding put too much restriction on the parameters, $LR = 34.59$, $df = 4$, $p = 0.33$. In other words, re-coding doesn't hurt the fit of the model (i.e., re-coding is fine).

I noted that the cross-level interaction between type of community and `(grpCmath)` is not significant (.ie., $t_{156} = -1.49$, $p = .14$, I dropped it from model. Also, note that LR test statistic = 2.22, $df = 1$, p=.14 for the cross-level interaction, which also confirms dropping cross-level and using re-coded type community is a reasonable decision.

(b) *Hours watching TV or videos:* I used a LR test to check whether hours watching TV/videos should be numeric or categorical (i.e., a factor) The $LR = 19.473$, $df = 3$, $< .01$ which indicates it should be a factor/cateogrical variables.

Based on my subjective views, I tested (using contrasts) TV=2 =3= 4 (n.s.) and then tested whether 2=3=4 versus 5 (significant). The latter was significant with $p < .01$. I recoded into 2 levels: some TV versus a lot of TV. I did a final test of whether even needed hours watching TV (significant, $p < .01$).
The re-coding hours watching TV

$$\text{TV} = \begin{cases} 1 & \text{if } \leq 4 \text{ hrs (some)} \\ 5 & \text{if } > 4 \text{ hrs (a lot)} \end{cases}$$

Using re-coded TV as.factor, the results are in table.

Table 1: Model refinements

Model	$-2\times$ LnLike	#	AIC	BIC.new	$\hat{\tau}$s	$\hat{\sigma}^2$	R_1^2	R_2^2
base	48255.4	18	48291.4	48399.5	$\tau_0^2 = 3.40$	50.90	.45	.80
					$\tau_1^2 = 0.18$	$\rho = .51$		
null	48261.3	16	48293.3	48395.4	$\tau_0^2 = 3.40$	51.08	.45	.80
re-code location	48260.0	14	48288.0	48368.6	$\tau_0^2 = 3.48$	50.90	.45	.79
					$\tau_1^2 = 0.20$	$\rho = .49$		
drop cross level	48262.2	13	48288.2	48362.00	$\tau_0^2 = 3.48$	50.90	.45	.79
					$\tau_1^2 = 0.21$	$\rho = .50$		
tv cat	48242.8	16	48274.8	48369.1	$\tau_0^2 = 3.41$	50.81	.45	.80
					$\tau_1^2 = 0.19$	$\rho = .52$		
re-code tv	48248.8	13	48274.8	48348.6	$\tau_0^2 = 3.41$	50.81	.45	.80
					$\tau_1^2 = 0.19$	$\rho = .44$		
games cat	48237.0	16	48269.0	48363.4	$\tau_0^2 = 3.33$	50.78	.46	.80
					$\tau_1^2 = 0.19$	$\rho = .50$		
re-code games	48240.0	13	48266.0	48339.7	$\tau_0^2 = 3.36$	50.79	.45	.80
					$\tau_1^2 = 0.19$	$\rho = .50$		
null for final	48246.2	11	48268.2	48336.0	$\tau_0^2 = 3.35$	50.89	.45	.80

(c) *Hours playing computer games.* I did similar steps for computer games as I did for hours watching TV; that is, test of all except 5 equal which was not significant ($p = .63$) and the test of 1=2=3=4 versus 5 is significant ($p < .01$). I recoded in a similar fashion as I did for hours TV.

Since the big distinction is between a lot versus some, I re-coded hours playing computer games as

$$\text{Computer Games} = \begin{cases} 0 & \text{if } \leq 4 \text{ hrs (some)} \\ 1 & \text{if } > 4 \text{ hrs (a lot)} \end{cases}$$

The results in terms of goodness-of-fit are reported in summary table.

I did a likelihood ratio test and found that the added restriction imposed by recoding was not significant: $LR = 6.09$, $df = 3$, $p = .11$.

5. *Starting with the base model from computer lab3, I refined this model to obtain my "best" model (i.e. simplify by dropping effects, re-coding a discrete variable, etc.).*

Summarize the steps that you and why you took them. This includes how you used the information from your contrasts, information criteria, tests of parameters, etc.

Using information from this homework and other analyses, I know the following:

- I need a random intercept.
- I need a random slope for `(grpCmath)`$_j$.
- I can simplify the model by re-coding type of community, hours watching TV (which is more categorical than numerical), and hours playing computer games (which is also more categorical than numerical). I found this from the contrasts.
- Below are refinements done successively to show that re-coding of community is fine, hours watching TV and playing computer games should be treated as categorical variables, and that proposed re-coding is fine. Lastly, I re-tested to see whether I still need a random slope (yes).

  Note: Comparing last model with and without random slope, the mixture was between χ^2_1 ($p = .012$) and χ^2 ($p = .04$), which lead to a $p = .03$).

- Although my final model doesn't have the smallest values on all the information criteria, I selected it because
  (a) It has small or almost the smallest values on the IC.
  (b) All of the effects in the model are significant.
  (c) It is parsimonious.
  (d) The interpretation makes sense.

6. *If your final model has a random slope, re-check to make sure that you need it. Report your results.*

   The null hypothesis is $H_o : \tau_1^2 = \tau_{01} = 0$. The test statistic equals 6.093

```
LR statistic df1 df0     p1     p0 pvalue
       6.243   2   1 0.044 0.012  0.028
```

   Comparing $\lambda = 6.24$ to the χ^2_1, it has a $p = .01$ and compared to χ^2_2 it has a $p = .05$, which gives our test a $p = .03$. Therefore, reject H_o, the data support that conclusion that we need a random slope.

7. *Give a full interpretation of the final model. Also give the HLM, linear mixed model and marginal model formulations using the parameter estimates.*

   Grading note: Everyone could have different final models based on decision made along the way.

Parameter Estimates from my final model:

	Final	
(Intercept)	18.34	$(5.75)^{**}$
xSchoolCenterMath	28.56	$(3.14)^{***}$
gendergirl	-1.30	$(0.17)^{***}$
grade4	0.86	$(0.19)^{***}$
tv2	-0.84	$(0.23)^{***}$
cg2	-1.52	$(0.37)^{***}$
xSchoolMeanMath	4.01	$(0.17)^{***}$
typecommunity33	-1.08	$(0.46)^{*}$
xSchoolCenterMath:xSchoolMeanMath	-0.71	$(0.09)^{***}$
AIC	48265.95	
BIC	48355.23	
Log Likelihood	-24119.98	
Num. obs.	7097	
Num. groups: idschool	146	
Var: idschool (Intercept)	3.35	
Var: idschool xSchoolCenterMath	0.19	
Cov: idschool (Intercept) xSchoolCenterMath	0.40	
Var: Residual	50.79	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

As an HLM:

Hierarchical model :

Level 1 :

$$\begin{aligned}
\widehat{(\texttt{science})}_{ij} &= \hat{\beta}_{0j} + \hat{\beta}_{1j}(\texttt{grpCmath})_{ij} + \hat{\beta}_{2j}\text{girl}_{ij} + \hat{\beta}_{3j}\text{fourth}_{ij} \\
&\quad + \hat{\beta}_{4j}(\texttt{hours_TV})_{ij} + \hat{\beta}_{5j}(\texttt{hours_computer_games})_{ij} \\
&\quad + \hat{\beta}_{6j}\text{rural}_j + \hat{\beta}_{7j}\text{sub\&urban}_j + R_{ij}
\end{aligned}$$

where $\hat{\sigma}^2 = 50.79$.

Level 2 :

$$\begin{aligned}
\hat{\beta}_{0j} &= 21.25 + 4.01(\texttt{grpMmath})_j - 1.08(\text{sub/urban})_j \\
\hat{\beta}_{1j} &= 28.57 - 0.71(\texttt{grpMmath})_j \\
\hat{\beta}_{2j} &= -1.29 \\
\hat{\beta}_{3j} &= 0.86 \\
\hat{\beta}_{4j} &= -0.84 \\
\hat{\beta}_{5j} &= -1.53 \\
\hat{\beta}_{6j} &= -3.04 \\
\hat{\beta}_{7j} &= -3.98
\end{aligned}$$

and

$$\hat{\boldsymbol{T}} = \begin{pmatrix} 3.30 & 0.39 \\ 0.39 & 0.19 \end{pmatrix}$$

Estimated Linear mixed model :

$$\begin{aligned}
(\texttt{science})_{ij} &= 18.34 + 28.56(\texttt{grpCmath})_{ij} - 1.30\text{girl}_{ij} + 0.86\text{fourth}_{ij} \\
&\quad -0.84(\texttt{hours_TV})_{ij} - 1.52(\texttt{hours_computer_games})_{ij} \\
&\quad +4.01(\texttt{grpMmath})_j - 1.08(\text{community})_j + \\
&\quad -0.71(\texttt{grpMmath})_j(\texttt{grpCmath})_{ij}
\end{aligned}$$

On average, a student's science score is expected to be

1.30 points higher for boys

0.86 points higher for 4th graders than 3rd graders

0.84 points higher for students who watch TV less than 4 hours per week.

1.52 points higher for students who play computer games less than 4 hours per week.

The effect of math scores on science is

$$\hat{\beta}_{1j} = 18.34 - 0.71\texttt{(grpMmath)}_j;$$

that is, for a 1 standard deviation increase of student math score relative to their peers, the student's science score is expected change by $18.34 - 0.71\texttt{(grpMmath)}_j$ points. The higher the average math scores in a school, the lower the effect of a student's relative standing. Note that a 1 unit change in both $\texttt{(grpMmath)}_j$ and $\texttt{(grpCmath)}_{ij}$ corresponds to a 1 standard deviation.

There are also differences between schools in terms of the intercepts. The overall level of the science scores is influenced by a school's location school(lower science scores in urban & suburban schools, 1.08 points lower) and the mean math scores of students at the school (level of science scores are .88 points higher).