2007 Summer Epi/Biostat Summer Institute

Multi-level Models

Homework

In this exercise you will be asked to interpret some results from multi-level models. You can type your responses either in this document or in a new document. You should email this assignment to Elizabeth (ejohnson@jhsph.edu) by no later than July 15th.

Part I: The lunch intervention

Scientific question: Does the lunch intervention impact cognitive ability?

The data consists of 4 measures of cognitive ability including:Raven's score (ravens), arithmetic score (arithmetic), Verbal meaning (vmeaning), and total digit span score (dstotal). Also included in the data are the following variables:

Lunch intervention (trt: 0=control, 1=calorie 2=meat= 3=milk)
Baseline age (age_at_time0),
Gender (1=boy 0=girl)
Baseline head circumference (head_circ)
Socioeconomic status score (ses)
Mother's reading ability (readtest)
Mother's writing ability (writetest)
Visit number (rn = 1,2,3,4,5 for weeks 1 through 5)

There were 12 schools that participated in the study. The intervention group was randomly assigned to the school. A variable number of students participated within each school. Each child was assessed at 5 times, once per week; at each occasion, the measures of cognition were recorded.

Denote the school by the index i, the student by the index j, and the visit/week by index k.

Let Y_ijk be the raven's cognition score for visit/week k (k = 1, 2, 3, 4, 5), from subject j (j = 1, …, n_i), from school i (i =1, 2, …, 12).

First we will present some summary information from the data.

The number of children participating within each school is displayed in the table below:

```
tab schoolid
```

schoolid	Freq.	Percent	Cum.
1	40	7.33	7.33
2	27	4.95	12.27
3	59	10.81	23.08
4	91	16.67	39.74
5	12	2.20	41.94
6	51	9.34	51.28
7	43	7.88	59.16
8	53	9.71	68.86
9	67	12.27	81.14
10	20	3.66	84.80
11	42	7.69	92.49
12	41	7.51	100.00
Total	546	100.00	

The table below displays the number of children in each of the intervention groups.

trt	Freq.	Percent	Cum.
control	127	23.26	23.26
calorie	146	26.74	50.00
meat	131	23.99	73.99
milk	142	26.01	100.00
Total	546	100.00	

The distribution of students by school and intervention group is displayed in the table below.

```
table schoolid trt
```

schoolid	trt control	calorie	meat	milk
1	40			
2			27	
3				59
4		91		
5		12		
6			51	
7		43		
8			53	
9	67			
10	20			
11				42
12				41

The mean raven's cognition scores by intervention group are displayed in the table below:

```
table trt, c(mean ravens sd ravens)
```

trt	mean(ravens)	sd(ravens)
control	18.4389	2.557517
calorie	18.1457	3.24382
meat	18.5301	3.041299
milk	17.9306	2.979153

1. Below you will find the results of an ordinary least squares linear regression for the raven's cognitive scores on the lunch intervention treatment. Specifically, we fit the following model:

$$\text{Ave(ravens score)} = b0 + b1*\text{calorie} + b2*\text{meat} + b3*\text{milk}$$

where the variables calorie, meat and milk are indicators of inclusion in each intervention group. Therefore, the control group is the reference and the mean score for the control group is represented by the intercept, b0. Note that Stata labels the intercept as "_cons". In one complete sentence interpret the regression coefficients that each compare the calorie, meat and milk groups to the control group,

respectively.

Model for Ravens cognitive score

ravens	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
calorie	-.2932296	.1651898	-1.78	0.076	-.6171467	.0306875
meat	.0911374	.1704044	0.53	0.593	-.243005	.4252798
milk	-.5083678	.1664867	-3.05	0.002	-.8348281	-.1819076
_cons	18.43894	.1209374	152.47	0.000	18.2018	18.67609

2. Next, we wish to fit a random intercept model for the raven's cognitive scores on the lunch intervention treatment taking into account all possible sources of variance in the data. Write out the model formula for this model. Your model should include three variance components. Be sure to include information regarding the distributions that you are assuming with variances defined. I got you started ……

   Y_ijk = b0 + b1*calorie_ijk + b2*meat_ijk + b3*milk_ijk + u_i + …..
   where u_i ~ Normal(0, tau^2), tau^2 is the heterogeneity in ravens cognitive scores across schools.

3. Below you will find the results from fitting the random intercept model for the raven's cognitive score.

ravens	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
calorie	-.2671385	.2804876	-0.95	0.341	-.8168841	.2826071
meat	.1233772	.2842285	0.43	0.664	-.4337005	.6804548
milk	-.5235633	.2759191	-1.90	0.058	-1.064355	.0172282
_cons	18.43929	.200607	91.92	0.000	18.0461	18.83247

Variance at level 1 **This is the lowest level variance (corresponding to ijk)**

6.5508953 (.20426682)

Variances and covariances of random effects

***level 2 (id) **This is the second level variance (corresponding to ij)**
    var(1): 2.2728217 (.22912251)
***level 3 (school) **This is the highest level variance (corresponding to i)**
    var(1): .02935327 (.05318119)

i. Interpret the results (both the regression coefficients and random intercept variance).

ii. Compare the results with those from OLS regression.

iii. What is the fraction of the variance that is due to within-subject variation?

iv. What is the fraction of the variance that is due to within-school but between-subject variation?

v. And what is the fraction of the variance that is due to between-school variation?

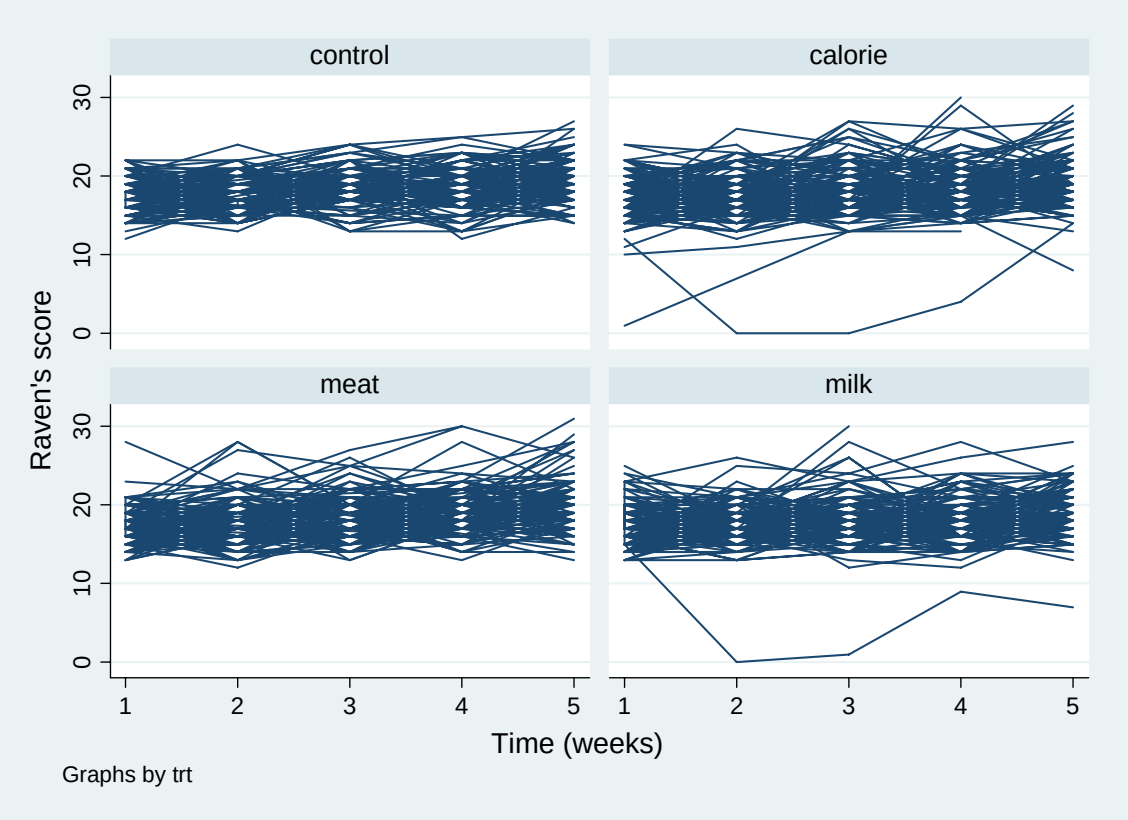
vi. Based on your calculation of the fraction of the different variance components, do you think it would be appropriate to simplify the model? Describe how you would simplify the model and also describe one graph/figure/table that you could have made to support your decision.

4. We ran the same analysis as in question 3 but further adjusting for baseline age, gender, baseline head circumference, socioeconomic status and mother's reading and writing ability. How do the results change after the adjustment for these relevant variables?

```
------------------------------------------------------------------------------
      ravens |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     calorie |  -.2770573   .2476244    -1.12   0.263    -.7623922    .2082775
     meat    |  -.0372297   .2524154    -0.15   0.883    -.5319548    .4574954
     milk    |   -.525485   .2519327    -2.09   0.037    -1.019264   -.0317059
age_at_time0 |   .1671888    .077657     2.15   0.031     .0149839    .3193938
      gender |    .301083   .1907292     1.58   0.114    -.0727393    .6749053
        ses1 |   .0043145   .0041504     1.04   0.299    -.0038202    .0124492
   head_circ |   .1899387    .066401     2.86   0.004     .0597951    .3200823
    readtest |   .0132151   .0295368     0.45   0.655    -.0446759    .0711061
   writetest |   .0242446   .0310859     0.78   0.435    -.0366826    .0851718
       _cons |   6.818578   3.340722     2.04   0.041     .2708823    13.36627
------------------------------------------------------------------------------
Variance at level 1
------------------------------------------------------------------------------
  6.3652302 (.21710862)
Variances and covariances of random effects
------------------------------------------------------------------------------
***level 2 (id)
    var(1): 1.9405232 (.22306077)
***level 3 (school)
     var(1): 1.081e-10 (.00001481)
------------------------------------------------------------------------------
```

5. Next we will study the longitudinal change in raven's score over time controlling for lunch intervention as well as baseline age, gender, baseline head circumference, socioeconomic status and mother's reading and writing ability.

The figure below displays the students' trajectories of raven's scores over time by intervention group.



NOTE: We will now ignore the index i since you established above that the degree of heterogeneity across schools was negligible. So let the index j now just count the total number of students and the index k still indicates the week of observation.

The linear random intercept model for this problem can be written out as follows:

$$Y_{jk} = b0 + b1*Time_{jk} + b2*calorie_j + b3*meat_j + b4*milk_j + b5*Z_j + u_j + e_{jk}$$

where Z_j contains all the adjustment variables, $u_j \sim Normal(0,tau^2)$ and $e_{ij} \sim Normal(0,sigma^2)$.

The results of fitting this model are presented below:

```
------------------------------------------------------------------------------
      ravens |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     calorie |  -.3026612   .2503153    -1.21   0.227    -.7932701    .1879477
     meat    |  -.0343622   .2562197    -0.13   0.893    -.5365436    .4678193
     milk    |  -.5552535   .2543817    -2.18   0.029    -1.053833   -.0566745
   rn (week) |   .5161691   .0367867    14.03   0.000     .4440686    .5882696
age_at_time0 |   .1676099   .0775365     2.16   0.031     .0156412    .3195786
      gender |   .2804431   .1914076     1.47   0.143    -.0947089    .6555951
        ses1 |   .0046623   .0041907     1.11   0.266    -.0035514     .012876
   head_circ |   .1970865   .0667083     2.95   0.003     .0663406    .3278324
    readtest |   .0106114   .0296023     0.36   0.720     -.047408    .0686309
   writetest |   .0303293   .0312139     0.97   0.331    -.0308488    .0915074
       _cons |   4.886603   3.340869     1.46   0.144    -1.661381    11.43459
------------------------------------------------------------------------------
Variance at level 1
------------------------------------------------------------------------------
  5.753903 (.19570873)
Variances and covariances of random effects
------------------------------------------------------------------------------
 ***level 2 (id)
   var(1): 2.1691922 (.24278962)
------------------------------------------------------------------------------
```

i. What type of correlation structure does this linear random effects model induce for the repeated measures within each subject?

ii. What is the estimate of the correlation of any two raven's scores taken from the same student?

iii. Interpret the slope for time (labeled as "rn (week)" in the Stata output).

6. Lastly, we fit a linear random intercept and random slope on the time variable. Starting with the model given in question 5, write out the model formula where we also want to allow the slope for time to vary across students. Be sure to define the covariance between the random intercept and random slope.

7. The results from fitting the random intercept and slope model are presented below.

ravens	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
calorie	-.2832644	.2461308	-1.15	0.250	-.7656718	.1991431
meat	-.1220936	.2518864	-0.48	0.628	-.6157819	.3715947
milk	-.5400549	.2504625	-2.16	0.031	-1.030952	-.0491574
rn (week)	.5163725	.0403634	12.79	0.000	.4372616	.5954834
age_at_time0	.1598898	.0772259	2.07	0.038	.0085299	.3112498
gender	.234042	.1901768	1.23	0.218	-.1386976	.6067816
ses1	.0037936	.0041198	0.92	0.357	-.0042811	.0118683
head_circ	.1831787	.0660044	2.78	0.006	.0538124	.312545
readtest	.0135618	.029353	0.46	0.644	-.043969	.0710927
writetest	.0261302	.030905	0.85	0.398	-.0344425	.0867029
_cons	5.77438	3.32495	1.74	0.082	-.7424034	12.29116

Variance at level 1 **This is the lowest level variance (corresponding to jk)**

5.2854362 (.20907831)

Variances and covariances of random effects

***level 2 (id)

var(1): 2.2831446 (.60330927) **This is the random intercept variance**

cov(2,1): -.2621916 (.16039095) cor(2,1): -.42572345 **This is the correlation between the subject specific random intercept and random slope.**

var(2): .16613034 (.0529639) **This is the random slope variance**

i. Interpret the slope for time from this model.

ii. What is the estimate of the variability in the within-in subject association between raven's scores and time? Using this information, we expect that 95% of all subjects slopes to fall within what range of the true slope?