Value Added Modeling

Dr. J. Kyle Roberts
Southern Methodist University
Simmons School of Education and Human Development
Department of Teaching and Learning

Background for VAMs

• Recall from previous lectures that we can define a growth model as:

\begin{align*}
Y_{ti} &= \pi_{0i} + \pi_{1i} \cdot t + e_{ti} \\
\pi_{0i} &= \beta_0 + u_{0i} \\
\pi_{1i} &= \beta_1 + u_{1i}
\end{align*}

• This is a two-level model with random intercepts and a random component for the time variable \( t \).

• This can also be viewed as a full model where:

\begin{align*}
Y_{ti} &= (\beta_0 + u_{0i}) + (\beta_1 + u_{1i}) \cdot t + e_{ti} \\
&= [\beta_0 + \beta_1 \cdot t] + [u_{0i} + u_{1i} \cdot t + e_{ti}] \\
\delta_i &\sim \mathcal{N}(0, \Omega), \quad e_{ti} \sim \mathcal{N}(0, \sigma^2 I)
\end{align*}

Setting Up The Model

• Before fitting the model we use two diagnostic plots to examine the data.

• First, we examine the functional form of the data using within-subject linear fits using OLS. This is accomplished using the \texttt{lmList} function.

• Second, we examine the variability of the intercepts and slopes to decide which random effects will be permitted to vary. This is accomplished using \texttt{confint}.

• The dataset we will be using is the \texttt{star} dataset in the \texttt{mlmRev} package and can be loaded simply by running \texttt{library(mlmRev)}.

Within-Subject Linear Fits

• Before fitting a linear model it is useful to examine the functional form of the data.

• Instead of plotting the entire dataset we take a random sample of 50 students (because our dataset is so large)

\begin{verbatim}
> samp <- sample(names(which(table(subset(star, + !is.na(math))$id) > 1)), 50)
> datsamp <- subset(star, id %in% samp)
> child.lm <- lmList(math ~ yrs | id, datsamp, na.action = na.omit)
\end{verbatim}

• The first two lines take the subset

• The last command fits the OLS regression and stores them in the object \texttt{child.lm.}
Within-Subject Confidence Intervals

- In some multilevel packages one typically fits the unconditional growth model and examines the chi-square statistic associated with the variance components. If the p-value is statistically significant, the random effects are retained.
- In R, this is accomplished a little differently. First, we examine the data graphically and we subsequently use a likelihood ratio test.
- Creating the visual displays requires use of the `confint` command.
- We have already created an object with the within-subject linear fits. We can now use the `plot` function to plot the intervals.
Fitting the VAM

• First, we will fit linear growth model with random effects for time.

\[ Y_{ti} = [\mu + \beta_1(year)] + [\delta_0 + \delta_1(year) + \epsilon_{ti}] \]

```r
> nograd <- list(gradient = FALSE, niterEM = 0)
> fm1 <- lmer(math ~ yrs + (yrs | id), star, control = nograd)
```

• The portion of code after the `control` statement turns off the analytic gradient and tells R to use no EM iterations. This is used to speed up computations when the number of grouping factors becomes exceedingly large.

Model Results

```r
> fm1
Linear mixed model fit by REML
Formula: math ~ yrs + (yrs | id)
Data: star
AIC BIC logLik deviance REMLdev
247325 247373 -123656 247311 247313
Random effects:
Groups   Name     Variance Std.Dev. Corr
id       (Intercept):1 1506.605 38.8150
          yrs:1       68.418  8.2715 -0.405
Residual            600.937 24.5140
Number of obs: 24613, groups: id, 10767

Fixed effects:
                  Estimate Std. Error t value
(Intercept)     484.2708   0.4947    978.9
yrs              43.2031   0.1962    220.2

Correlation of Fixed Effects:
                            (Intr)
yrs -0.612
```

Model Modifications

• We can now test our original random effects model against a model in which we do not allow for the time variable (`yrs`) to have random effects

```r
> fm2 <- lmer(math ~ yrs + (1 | id), star, control = nograd)
> anova(fm1, fm2)
```

Checking Assumptions

• The distributional assumptions underlying `fm1` are that the within-group residuals are distributed as \( \epsilon_{ti} \sim N(0, \sigma^2 I) \) and the random effects are distributed as \( \delta_i \sim N(0, \Omega) \).

• Both of these assumptions must be examined.
> with(fm1, xyplot(resid(.) ~ fitted(.) | gr, type = c("g", "+ "p")))

> qqmath(~resid(fm1))

> qqmath(~ranef(fm1)$id$'Intercept', cex = 0.7)

> qqmath(~ranef(fm1)$id$yrs, cex = 0.7)
Extending the Model

- The Project Star dataframe includes unique school IDs so we can extend this to a three-level model and examine the school effects.
- We previously noted how the structure of the random effects can be modified to handle additional "levels". R can handle an arbitrary number of levels when fitting models with nested random effects.
- The portion of code `(yrs|sch:id)` denotes that students are nested within schools. We could also run this portion of the code as `(yrs|id)+(yrs|sch)`, such that we have defined yrs to have random effects for both individuals and schools.

Extending the Model (cont.)

- We specify the model as:

  \[ Y_{ti} = [\mu + \beta_1(time)] + [\theta_{0j(i)} + \theta_{1j(i)}(time) + \delta_{0i} + \delta_{1i}(time) + \epsilon_{ti}] \]

  \[ \theta_j \sim N(0, \Phi), \quad \delta_i \sim N(0, \Omega), \quad \epsilon_{ti} \sim N(0, \sigma^2 I) \]

Examining the Variance Components

- As we did before, we use the `lmList` function to produce within-school confidence intervals.
- The only difference here is that we have 80 schools, so we do not take a subset.

  > school.lm <- lmList(math ~ yrs | sch, data = star)
> plot(confint(school.lm))

Model Testing

- Suppose we wanted to modify the structure of the random effects such that we did not model random slopes for `yrs`.

\[ Y_{it} = [\mu + \beta_1 \cdot time] + [\theta_{0j(i)} + \theta_{1j(i)} + \delta_{0i} + \epsilon_{it}] \]

> sch2.lmer <- lmer(math ~ yrs + (1 | id) + (1 | sch), star, control = nograd)
> anova(sch.lmer, sch2.lmer)

Data: star
Models:
sch2.lmer: math ~ yrs + (1 | id) + (1 | sch)
sch.lmer: math ~ yrs + (yrs | id) + (yrs | sch)
   Df AIC  BIC logLik deviance REMLdev
sch2.lmer 5 246225 246265 -123107
sch.lmer 9 244624 244697 -122303 1609.2 4 < 2.2e-16

Looking at Effects

> head(ranef(sch.lmer)$sch)

<table>
<thead>
<tr>
<th></th>
<th>(Intercept)</th>
<th>yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.003624</td>
<td>-4.254397</td>
</tr>
<tr>
<td>2</td>
<td>-30.828722</td>
<td>4.528853</td>
</tr>
<tr>
<td>3</td>
<td>16.818051</td>
<td>-5.660053</td>
</tr>
<tr>
<td>4</td>
<td>-9.848673</td>
<td>9.203151</td>
</tr>
<tr>
<td>5</td>
<td>-17.785526</td>
<td>0.764398</td>
</tr>
<tr>
<td>6</td>
<td>-15.086948</td>
<td>3.816181</td>
</tr>
</tbody>
</table>

> head(ranef(sch.lmer)$id)

<table>
<thead>
<tr>
<th></th>
<th>(Intercept)</th>
<th>yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>100017</td>
<td>93.515890</td>
<td>-4.107405381</td>
</tr>
<tr>
<td>100028</td>
<td>-40.9552884</td>
<td>1.798833869</td>
</tr>
<tr>
<td>100045</td>
<td>10.4070394</td>
<td>-0.01423357</td>
</tr>
<tr>
<td>100064</td>
<td>-0.1676666</td>
<td>0.007364254</td>
</tr>
<tr>
<td>100070</td>
<td>49.1877278</td>
<td>-1.999841478</td>
</tr>
<tr>
<td>100096</td>
<td>30.7611035</td>
<td>0.21232219</td>
</tr>
</tbody>
</table>