Stat 414 - Day 16
Longitudinal data, cont. (Ch. 15)

Last Time: Longitudinal data: Have repeat observations over time

- wide vs. long format
- time varying vs. time invariant variables
- explore the raw data (graphs, correlation matrix)
- *time* is often the only Level 1 variable
  - Consider how parameterized, what "0" represents, start at zero?
  - Consider form of association (e.g., linear, quadratic, piecewise)
  - Consider random slopes for time (models unequal variances, correlations)
- unconditional growth model: $y_{ij} = \beta_{0j} + \beta_{1j}time_{ij} + \epsilon_{ij}$
  - Random slopes for time $V(Y_{ij}) = \tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2$
  - Assumes (conditional) residuals on the same individual are independent of each other $Cov(Y_{ij}, Y_{kj}) = \tau_0^2 + \tau_{01}(x_{ij} + x_{kj}) + \tau_1^2(x_{ij}x_{kj})$

Example: Data were collected by the Minnesota Department of Education for all Minnesota schools during the years 2008-2010 to compare charter and non-charter schools. Does the model match the data?

```
cor(matrix, use="pairwise.complete.obs")
##                 MathAvgScore.0 MathAvgScore.1 MathAvgScore.2
## MathAvgScore.0      1.0000000      0.8064146      0.7727215
## MathAvgScore.1      0.8064146      1.0000000      0.8331408
## MathAvgScore.2      0.7727215      0.8331408      1.0000000
```

raw data

time w/ random slopes

For the unconditional growth model, compare the estimated response variances and the correlation matrix to the raw data.

Conditional variance-covariance	Marginal variance-covariance

`## 1 2 3`
`## 1 8.8202 0.0000 0.0000`
`## 2 0.0000 8.8202 0.0000`
`## 3 0.0000 0.0000 8.8202`
	`## 1 2 3`
	`## 1 48.263 40.952 42.462`
	`## 2 40.952 51.39 44.192`
	`## 3 42.462 44.192 54.742`

Conditional correlations	**Marginal correlations**

`## 1 2 3`
`## 1 1 0 0`
`## 2 0 1 0`
`## 3 0 0 1`
	`## 1 2 3`
	`## 1 1.0000000 0.8222865 0.8261001`
	`## 2 0.8222865 1.0000000 0.8331700`
	`## 3 0.8261001 0.8331700 1.0000000`

$\hat{Var}(e_{11})$ $\hat{Var}(e_{21})$ for any one school $\hat{\sigma}^2$

$Var(Y_{1i}) = Var(Y_{1i})$ ... models predictions for $\hat{Var}(Y_{21}) = \hat{Var}(Y_{22}) = \cdots$

(a) Compare and contrast the correlation structures between the raw data and the model.

raw data has unequal correlations (which 2 time points) but with "decay" that the model is not capturing

AR(1) Errors So far we have assumed that the "occasion-specific" residuals (the ϵ's) are independent: $cov(\epsilon_{ij}, \epsilon_{kj})$ = 0 for any pair of occasions on the same individual.

A common alternative covariance structure is an AR(1) model for the Level 1 residuals, which assumes the covariance matrix of the errors is of the form

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & & & \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \\ \\ \end{array} & \sigma^2 \begin{bmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho^2 & \rho & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix} \end{array}$$

(b) What does the model assume for $Var(\epsilon_{ij})$?

$Var(\epsilon_{ij}) = \sigma^2$

(c) What does the model assume for $cov(\epsilon_{ij}, \epsilon_{kj})$? $corr(\epsilon_{ij}, \epsilon_{kj})$? How do these change the further apart the measurements in time?

$\epsilon_{ij} = \rho\epsilon_{(i-1)j} + \xi_{ij}$ where $\xi_{ij} \sim N(0, \sigma^2)$

$var(\epsilon_{ij} | \epsilon_{(i-1),j}) = \sigma^2$

$cov(\epsilon_{ij}, \epsilon_{kj}) = \rho^{|i-k|}\sigma^2$, $corr(\epsilon_{ij}, \epsilon_{(i-1)j}) = \rho^{|i-k|}$

There is a "serial dependence" in the residuals but otherwise uncorrelated and constant variance, but gives us the "decay" in the correlation

(d) Derive the expression for $cov(y_{ij}, y_{kj})$ for the AR(1) model.

(No random slopes)

- $cov(y_{ij}, y_{kj})$
  $= cov(u_{oj} + \epsilon_{ij}, u_{oj} + \epsilon_{kj})$
  $= cov(u_{oj}, u_{oj}) + cov(\epsilon_{ij}, \epsilon_{kj})$ No cross-level correlation in random effects
  $= \tau_0^2 + \rho^{|i-k|}\sigma^2$

(f) How many additional parameters does this add to our model? 1

So instead of random slopes on time, fit the AR(1) structure.

```
model2 = lme(MathAvgScore ~ year08 + I(year08^2), random = ~1 | schoolnum,
correlation=corAR1(), data = chart_long); summary(model2)
Random effects:
 Formula: ~1 | schoolnum
        (Intercept) Residual
StdDev:    6.464088  3.08765
```

τ σ

```
Correlation Structure: AR(1)
 Formula: ~1 | schoolnum
 Parameter estimate(s):
```

~~Phi~~ Rho

```
0.1418763
```

$\hat{\rho}$

```
 Correlation structure:
```

CI for ρ

```
          lower      est.     upper
Phi -0.05885505 0.1418763 0.3315805
```

```
 Within-group standard error:
```

~~error~~ dev, CI for σ

```
   lower     est.    upper
2.791967 3.087650 3.414646
```

(g) What is the estimated parameter of the AR(1) model ("autocorrelation"). How do you interpret it? Is it statistically significant? How are you deciding?

$\hat{\rho} = .142$ estimated conditional correlation for Level 1 residuals one time point apart

not significant because 0 is in CI for ρ

```
## Random effects:
##  Formula: ~1 | schoolnum
##         (Intercept) Residual
## StdDev:    6.464088  3.08765
```

τ_0 σ

Conditional variance-covariance

```
##           1        2         3
## 1 9.5335795 1.352589 0.1919003
## 2 1.3525888 9.533580 1.3525888
## 3 0.1919003 1.352589 9.5335795
```

Marginal variance-covariance

```
## Marginal variance covariance matrix
##        1      2      3
## 1 51.318 43.137 41.976
## 2 43.137 51.318 43.137
## 3 41.976 43.137 51.318
```

Conditional correlations

```
##            1         2          3
## 1 1.00000000 0.1418763 0.02012888
## 2 0.14187628 1.0000000 0.14187628
## 3 0.02012888 0.1418763 1.00000000
```

Marginal correlations

```
##           1         2         3
## 1 1.0000000 0.8405825 0.8179649
## 2 0.8405825 1.0000000 0.8405825
## 3 0.8179649 0.8405825 1.0000000
```

(h) Show how to find the correlation between year 1 and year 3 residuals based on the correlation between year 1 and year 2 residuals.

years 1 & 2 : .142

years 1 & 3 : $.142^2 = .0201$

calc problem

(i) Show how to find the "marginal" variance at time 0. What about time 1 and time 2?

(j) Show how to find the values in the marginal variance-covariance and correlation matrices

(k) Does the correlation matrix appear to be a better fit to the data?

yes

Notes:

- The AR structure does assume the observations are equally spaced in time (e.g., one year to the next/same distance apart) for all individuals. The AR model also assumes the variance is the same at the different time points, just allows for this consistent drop off in correlation as time points are further apart.
- There are more flexible structures, but “in many applications, AR(1) provides an adequate model of the within subject correlation, providing more power without sacrificing Type I error control.”
- From Roback and Legler (2019): In the charter school example, as is often true in multilevel models, the choice of covariance matrix does not greatly affect estimates of fixed effects. The choice of covariance structure could potentially impact the standard errors of fixed effects, and thus the associated test statistics, but the impact appears minimal in this particular case study. In fact, the standard model typically works very well. So is it worth the time and effort to accurately model the covariance structure? If primary interest is in inference regarding fixed effects, and if the standard errors for the fixed effects appear robust to choice of covariance structure, then extensive time spent modeling the covariance structure is not advised. However, if researchers are interested in predicted random effects and estimated variance components in addition to estimated fixed effects, then choice of covariance structure can make a big difference. For instance, if researchers are interested in drawing conclusions about particular schools rather than charter schools in general, they may more carefully model the covariance structure in this study.