

finding your inner ancova

Linear Mixed Models

Stat 209

Appendix to An R and S-PLUS Companion to Applied Regression

John Fox

3.2 Fitting a Hierarchical Linear Model with lme (HLM)

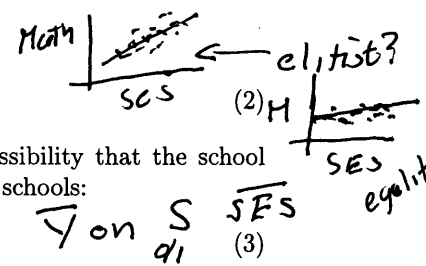
Following Bryk and Raudenbush (1992) and Singer (1998), I will fit a hierarchical linear model to the math-achievement data. This model consists of two equations: First, within schools, we have the regression of math achievement on the individual-level covariate SES; it aids interpretability of the regression coefficients to center SES at the school average; then the intercept for each school estimates the average level of math

160 schools
7185 students

achievement in the school.

Using centered SES, the individual-level equation for individual j in school i is

$$\text{mathach}_{ij} = \alpha_{0i} + \alpha_{1i}\text{cse}_{ij} + \epsilon_{ij}$$



At the school level, also following Bryk and Raudenbush, I will entertain the possibility that the school intercepts and slopes depend upon sector and upon the average level of SES in the schools:

ancova \bar{y}
 $\hat{\beta}_{01} = 5.3$
 $\hat{\beta}_{02} = 1.22$
 ancova on slopes, Math/SES
 $\hat{\beta}_{11} = 1.04$
 $\hat{\beta}_{12} = -1.64$

1) ancova on class means

$$\alpha_{0i} = \gamma_{00} + \gamma_{01}\text{meanses}_i + \gamma_{02}\text{sector}_i + u_{0i}$$

2) ancova on slopes, Math/SES

$$\alpha_{1i} = \gamma_{10} + \gamma_{11}\text{meanses}_i + \gamma_{12}\text{sector}_i + u_{1i}$$

This kind of formulation is sometimes called a coefficients-as-outcomes model.

Substituting the school-level equation 3 into the individual-level equation 2 produces

$$\text{mathach}_{ij} = \gamma_{00} + \gamma_{01}\text{meanses}_i + \gamma_{02}\text{sector}_i + u_{0i} + (\gamma_{10} + \gamma_{11}\text{meanses}_i + \gamma_{12}\text{sector}_i + u_{1i})\text{cse}_{ij} + \epsilon_{ij}$$

Rearranging terms,

$$\text{mathach}_{ij} = \gamma_{00} + \gamma_{01}\text{meanses}_i + \gamma_{02}\text{sector}_i + \gamma_{10}\text{cse}_{ij} + \gamma_{11}\text{meanses}_i\text{cse}_{ij} + \gamma_{12}\text{sector}_i\text{cse}_{ij} + u_{0i} + u_{1i}\text{cse}_{ij} + \epsilon_{ij}$$

Catholic schools higher achievement more egalitarian (compensatory)

combined model eq. Berk sec 10.3 (10.15)

Here, the γ 's are fixed effects, while the u 's (and the individual-level errors ϵ_{ij}) are random effects.

Finally, rewriting the model in the notation of the linear mixed model (equation 1),

$$\text{mathach}_{ij} = \beta_1 + \beta_2\text{meanses}_i + \beta_3\text{sector}_i + \beta_4\text{cse}_{ij} + \beta_5\text{meanses}_i\text{cse}_{ij} + \beta_6\text{sector}_i\text{cse}_{ij} + b_{i1} + b_{i2}\text{cse}_{ij} + \epsilon_{ij}$$

even though there are interactions galore in combined model, none in the Level 2 ancova models

lme (linear mixed effects) function in the nlme library, however, employs the Laird-Ware form of the linear mixed model (after a seminal paper on the topic published by Laird and Ware, 1982):

$$y_{ij} = \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + b_{i1} z_{1ij} + \dots + b_{iq} z_{qij} + \epsilon_{ij} \quad (1)$$

$$b_{ik} \sim N(0, \psi_k^2), \text{Cov}(b_k, b_{k'}) = \psi_{kk'}$$

$$\epsilon_{ij} \sim N(0, \sigma^2 \lambda_{ijj}), \text{Cov}(\epsilon_{ij}, \epsilon_{ij'}) = \sigma^2 \lambda_{ijj'}$$

where

- y_{ij} is the value of the response variable for the j th of n_i observations in the i th of M groups or clusters.
- β_1, \dots, β_p are the fixed-effect coefficients, which are identical for all groups.
- x_{1ij}, \dots, x_{pij} are the fixed-effect regressors for observation j in group i ; the first regressor is usually for the constant, $x_{1ij} = 1$.
- b_{i1}, \dots, b_{iq} are the random-effect coefficients for group i , assumed to be multivariately normally distributed. The random effects, therefore, vary by group. The b_{ik} are thought of as random variables, not as parameters, and are similar in this respect to the errors ϵ_{ij} .

Mixed effects
 $y = X\beta + Zb + \epsilon$

HSB by ancova

Week 5 Stat 209
P.2

```
> #let's do the hsb ancova
> hsbancdat = read.table(file="D:\\drr09\\stat209\\hsbancova", header = T)
> summary(hsbancdat)
  school      Intercept      csesslp      sector      meanses
  20          21          21          PIC
```

```
> attach(hsbancdat)
> tapply(Intercept, sector, summary)
$C  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   7.336  13.200  14.470  14.200  15.900  19.720
$P  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   4.240   9.719  11.710  11.390  13.200  18.110
> tapply(csesslp, sector, summary)
$C  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -2.0150  0.5698  1.5230  1.4680  2.4600  5.2580
$P  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -1.014   1.695   2.922   2.772   3.824   6.266
```

Cath higher on \bar{Y}

Cath lower on $\hat{\alpha}_1$ school slope

```
> #initial differences on covariate?
> tapply(meanses, sector, summary)
$C  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -0.7619 -0.1039  0.2388  0.1601  0.4346  0.8250
$P  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -1.19400 -0.40090 -0.09058 -0.13550  0.11370  0.68200
> #Cath higher on school-level SES
```

so ancova will adjust a little bit

```
> # do the ancova on school-level outcomes (level, slope)
> hsbancdat$gr = 2 - as.numeric(sector) # code Cath = 1, Pub = 0 on sector
> intancova = lm(Intercept ~ gr + meanses)
```

make 0,1 group var

```
> summary(intancova)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.1195    0.2026  59.807 < 2e-16 ***
gr           1.2219    0.3169   3.855 0.000168 ***
meanses     5.3874    0.3810  14.140 < 2e-16 ***
```

school level close

alternative: individual level ancova? $Y_{ij}, G, \text{ses}_{ij}$

```
---
Residual standard error: 1.859 on 157 degrees of freedom
Multiple R-squared: 0.6489, Adjusted R-squared: 0.6444
F-statistic: 145.1 on 2 and 157 DF, p-value: < 2.2e-16
```

```
> # compare with hlm/lme coeffs gr vs 1.226 (.306) df 157 t = 4.00
> # compare with hlm/lme coeffs meanses vs 5.33 (.369) df 157 t = 14.4
```

```
> slpancova = lm(csesslp ~ gr + meanses)
> summary(slpancova)
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.8886    0.1600  18.049 < 2e-16 ***
gr          -1.5580    0.2503  -6.224 4.22e-09 ***
meanses     0.8612    0.3009   2.862 0.00478 **
```

school slope

```
---
Residual standard error: 1.468 on 157 degrees of freedom
Multiple R-squared: 0.1999, Adjusted R-squared: 0.1897
F-statistic: 19.61 on 2 and 157 DF, p-value: 2.492e-08
```

```
> # compare with hlm/lme coeffs gr vs -1.64 (.239) t = -6.85
> # compare with hlm/lme coeffs meanses vs 1.03 (.299) df 157 t = 3.48
```

a little different, due to weighting.

```
> # look at comparing regressions; neither outcome refutes parallel regressions
> intcnrl = lm(Intercept ~ gr + meanses + I(gr*meanses))
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.1825    0.2075  58.714 < 2e-16 ***
gr           1.2487    0.3168   3.942 0.000122 ***
meanses     5.8524    0.5139  11.388 < 2e-16 ***
I(gr * meanses) -1.0261    0.7634  -1.344 0.180855
```

dependence of sector effect on SES

```
> slpcnrl = lm(csesslp ~ gr + meanses + I(gr*meanses))
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.9122    0.1646  17.692 < 2e-16 ***
gr          -1.5480    0.2513  -6.160 5.92e-09 ***
meanses     1.0351    0.4077   2.539 0.0121 *
I(gr * meanses) -0.3837    0.6056  -0.634 0.5272
```