3.

I. Questions about anything?

■ One-Way ANOVA (fully unconditional or null) model 복습

Model:

Level-1 model: $Y_{ij} = \beta_{0j} + r_{ij}$, $r_{ij} \sim N(0,\sigma^2)$ σ^2 = **level-1 variance**

Level-2 model: $\beta_{0j} = \gamma_{00} + u_{0j}$, $u_{0j} \sim N(0,\tau_{00})$ τ_{00} = **level-2 variance**

β_{0j} = mean outcome for unit j

γ_{00} = grand-mean outcome in the population (우리의 모집단은 L−1!)

u_{0j} = random (group) effect associated with unit j (L-2 효과)

r_{ij} = level-1 error term

HLM 분석을 시작하면서 점검하게 되는 모형인데 이 모형을 통해

(1) Estimate grand mean for outcome (dependent) variable
(2) Estimate variance components at both levels
(3) Estimate intra-class corrections and reliability
(4) Test hypothesis whether all j units have same mean

- **이 모형은 우리가 average or expected outcomes among schools (level-2 units)이외에 관심이 없을 때는 유용하다.**
- **하지만 "L−1 독립변수들"이 "종속변수와 L−2"에 상관되어 있는데 L−1 의 독립변수를 모형에 포함하지 않으면 L−2 의 효과를 편파적으로 추정하게 된다. → ANCOVA 상황!**

■ **One-Way ANCOVA Model (ANCOVA?)**

•ANOVA 는 우리가 관심을 가진 종속변수(e.g. 임금) 의 평균값이 집단간 차이(예: 여자 vs. 남자)가 있는지를 비교한다. 하지만 많은 경우 우리는 여자와 남자의 임금을 비교하면서 다른 양적 변수(covariate)를 통제한다. 이러한 경우가 ANCOVA!

- 예를 들어 여자와 남자 사이에 교육의 차이나 일 경험의 차이를 통제해야만 한다. 대개의 경우 이러한 변수들은 종속변수에 영향을 주면서 동시에 독립변수에 해당하는 집단(여자 vs 남자) 변수에도 상관되어 있다.
- 간단히 특정 집단은 다른 집단에 비해 covariate 의 평균 점수 높다. 이러한 상황에서 ANCOVA 분석의 결과는 ANOVA 분석의 결과와 다르게 된다.
- 정리하면, 교육연한이나 일 경험은 임금수준에 긍정적 효과를 주는 것으로 알려져 있는데 특정 업무를 하는 사람들 중에서 남자들이 교육연한이나 일 경험에 대한 변수 값이 더 높을 때 그 업무에 종사하는 여자와 남자의 평균임금수준에 대한 비교분석 결과는 교육연한이나 일 경험을 통제했는지 여부에 달려있게 될 것이다.

■학교수준에서 학생들의 특성(위에서 말한 covariate, e.g. SES)이 차이를 보일 수 있다(특목고 vs. 일반고).

■**Thus we want to know expected outcomes for schools if they enrolled similar populations or expected outcomes for same type of student**
--e.g., students of similar SES backgrounds
--e.g., students of same ethnicity of race

➢ (좋은 학생을 뽑는 학교? 잘 가르치는 학교?)

➢ 학부모들은 진심으로 아래 질문에 관심이 있나? (^^;)

✧**What difference does the school a child goes to make in the child's achievement?**

■ **HLM** 에서 이러한 질문에 답하기 위해서는 **level-1** 독립변수(**predictors**)를 도입해야 한다.

Model:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$, $r_i \sim N(0,\sigma^2)$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$, $u_{0j} \sim N(0,\tau_{00})$

$\beta_{1j} = \gamma_{10}$ **[no random term]** (무슨 뜻 ?)

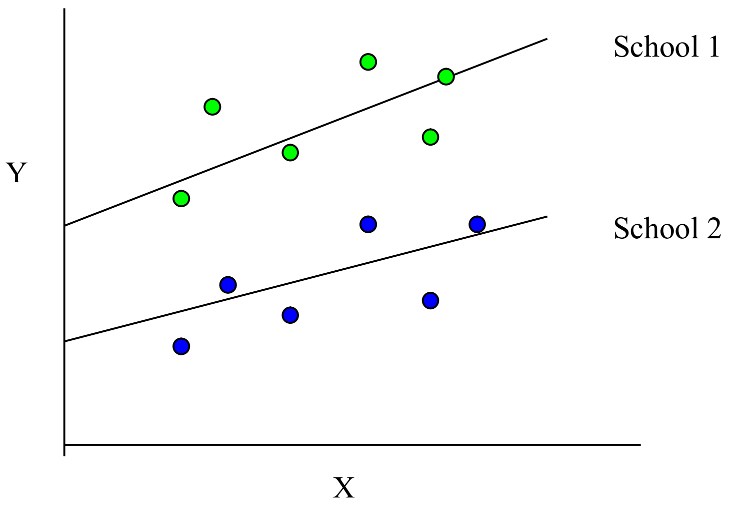
Fixed effects: γ_{00} = **average outcome for sample of groups**

γ_{10} = **average individual effect (slope) on outcome**

Random effects: u_{0j} = **unique effect of group j on average outcome**

- **그래서 level-1 predictor(s) 도입하려면, 우리는 아래와 같은 문제에 대해 결정해야 한다.**
    1) **Whether to introduce random coefficients**
    2) **Whether to center or transform the level-1 predictor**

1. Fixed or random slopes?



❑통상, 이 문제는 **empirically** 정할 수 있다:

- ◆ **Assume random slopes and test hypothesis:**
  H_0: τ_{11}= 0 (τ_{11} chi-squared distribution with J-2 DOF)

❑하지만 어떤 경우는 **slope** 들은 분석의 출발부터 **Fixed** 할 수 있다.

Why?

1. **Not interested in between-school differences in slopes**
2. **Small within-school samples may prevent estimating slopes reliably**

2. Centering

■ **the meaning or interpretation of the intercept, β_{0j}**

a. **Non-centered**

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$$

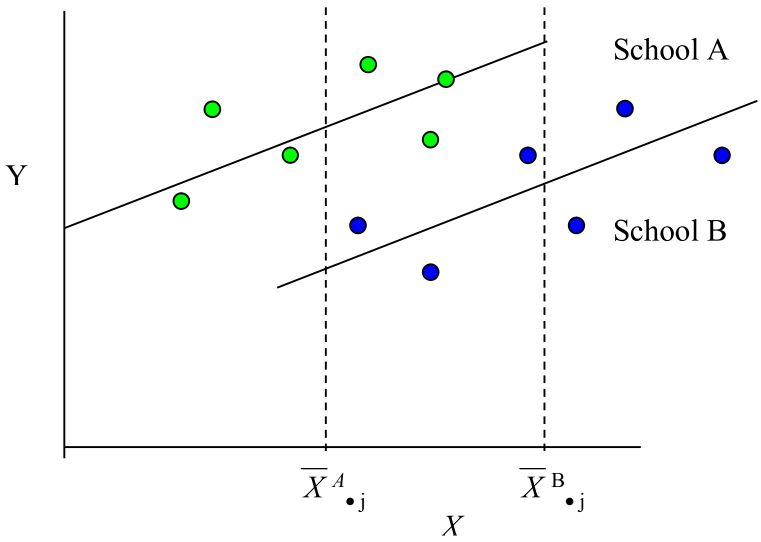
β_{0j} = expected value of Y_{ij} for student i in school j when X_{ij} = 0.

문제는 독립변수 X_{ij} 가 현실적 의미가 있는지?

예: **X_{ij} = previous achievement level,**

하지만 **previous achievement level=0** 은 현실적이지 않다.

b. Group-mean centered



$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(X_{ij} - \overline{X}_{\bullet j}) + r_{ij}$$

$\beta_{0j} = \mu_{Yj}$ or unadjusted mean for school j

= expected value of Y_{ij} when $X_{ij} = \overline{X}_{\bullet j}$

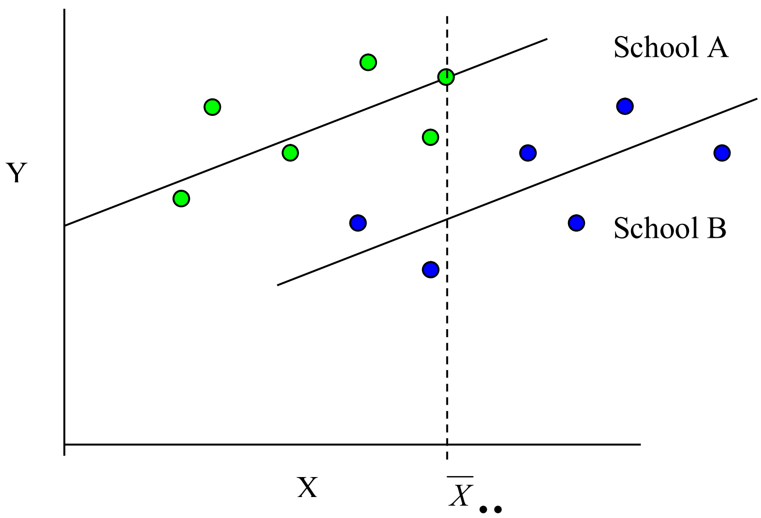
(mean value for school j),

i.e. for the "average student" in school j

Var (β_{0j}) is variance among level-2 units

Problem: $X_{ij} = \overline{X}_{\bullet j}$ 일때 Y_{ij} 기대치는 β_{0j} (절편) 이다. 그런데 학교 A 의 $\overline{X}_{\bullet j}$ 와 학교 B 의 $\overline{X}_{\bullet j}$ 값이 다르기 때문에 We can't compare intercepts.

c. Grand-mean centered



- $Y_{ij} = \beta_{0j} + \beta_{1j}\,(X_{ij} - \overline{X}_{\bullet\bullet}) + r_{ij}$

- **$\beta_{0j} = \mu_{Yj} - \beta_{1j}\,(\overline{X}_{\cdot j} - \overline{X}_{\bullet\bullet})$ or adjusted mean for school *j***
  = expected value of Y_{ij} when $X_{ij} = \overline{X}_{\bullet\bullet}$ (grand mean),
  i.e. for the "average student" in the sample of schools
- **Var (β_{0j}) = τ_{00} variance among schools in adjusted means**
- Grand-mean centering 에서는 $X_{ij} = \overline{X}_{\bullet\bullet}$ 일 때, Y_{ij} 기대치(predicted value) β_{0j} (절편)이다. 그런데 학교 A 에서도 학교 B 에서도 $X_{ij} = \overline{X}_{\bullet\bullet}$인 값은 같다. We can compare intercepts(β_{0j})

d. Case of dummy variables:

$Y_{ij} = \beta_{0j} + \beta_{1j}\,X_{ij} + r_{ij}$

Example Xij= FEMALE

❑ **Group-mean centering:**

◆ **$X_{ij} - \overline{X}_{\bullet j}$ represents proportion of males in the school (왜냐하면 $\overline{X}_{\bullet j}$ proportion of female) and β_{0j} still represents average outcome for school *j***

❑ **Grand-mean centering:**

$X_{ij} - \overline{X}_{\bullet\bullet}$ represents proportion of males in the population and

β_{0j} represents adjusted mean, adjusted for differences in proportion of females

Example:

ANCOVA 에선 Grand Mean Centering! 왜? 위에서 설명한 바와 같이 covariate 는 Level-1 종속변수와 영향을 주지만 Level-2 를 구성하는 집단(학교)에 따라 개인들의 covariate 값이 다를 수 있기 때문에 (학교수준에서 해당 covariate 평균이 다른 경우 조정, adjust!)

Model 1. one way ANCOVA with random intercept

LEVEL 1 MODEL

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

MIXED MODEL

$$MATHACH_{ij} = \gamma_{00} + \gamma_{10} * SES_{ij} + u_{0j} + r_{ij}$$

앞선 시간에 말했듯이 표준화 되어 있어서 우리 실습자료는 센터링을 안 해도 자동적으로 grand-mean centering 효과!

```
Sigma_squared =      37.03440
 Tau
 INTRCPT1,B0          4.76815
Tau (as correlations)
 INTRCPT1,B0  1.000
 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                          0.843
 ----------------------------------------------------

The value of the likelihood function at iteration 6 = -2.332167E+004
```

```
 The outcome variable is  MATHACH
 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                            Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For          INTRCPT1, B0
    INTRCPT2, G00           12.657481    0.187984     67.333       159     0.000
 For         SES slope, B1
    INTRCPT2, G10            2.390199    0.105719     22.609      7183     0.000
 ----------------------------------------------------------------------------
```

❑Note that we now estimate two fixed effects – the intercept and the effect of student's SES.
❑The intercept γ_{00} is no longer the average math achievement – it is now math achievement for someone with all predictors equal to zero.

1. In this case, it's math achievement for someone with SES=0, but because the SES scale was designed to have a mean of 0, the intercept (12.66) is essentially the math achievement for someone with average SES.

(즉, 자동적으로 grand mean centered, 그래서 centering 불필요)

2. The effect of SES, γ_{10}, can be interpreted as follows: one unit increase in SES is associated with 2.39 unit increase in one's math achievement. So math achievement for someone with SES being 1 unit above the mean would be: 12.66+2.39=15.05

❑Note that each β_{0j} is now ***the mean outcome for each group (i.e. school) adjusted for the differences among these groups in SES***.

```
 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect            Standard      Variance     df    Chi-square  P-value
                          Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        2.18361       4.76815   159    1037.09077    0.000
  level-1,       R         6.08559      37.03440
 -----------------------------------------------------------------------------

 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                         = 46643.331427
 Number of estimated parameters = 2
```

❑As we now accounted for some portion of the variance by controlling for SES, we can calculate the adjusted intra-class correlation: ρ=4.76815/(4.76815+37.03440)= .11406362
❑The decrease in ρ from .18035673 to .11406362 reflects a reduction in the relative share of between-school variance (correlation between pairs belonging to the same school) when we control for student SES. But there is still significant variation across schools.
❑We could also calculate the proportion of variance explained at each level by comparing the current variance estimates to those in the null model (fully unconditional model). (This is the easiest method recommended by Bryk and Raudenbush; another method is p.74)

- **between-school:** (8.61431 - 4.76815)/8.61431 = .44648498

- **within-school variance:** (39.14831 - 37.03440)/ 39.14831 = .05399748

❑ **So, controlling for individuals' SES explained 45% of between-school variance, and 5% of within-school variance in math achievement.**

❑ **We could also calculate the total percentage of variance explained:**

■ **(39.14831+8.61431-37.03440-4.76815)/(39.14831+8.61431)= .12478524**

◆**So students' SES explained 12% of the total variance in math achievement.**

◆**Some researcher report this as R^2 of Level-1. (We will review such an article together if time allows)**

❑**이 분석을 자세히 다루는 이유 (교재 p.142 참고, 다른 사례 p.112):**

1. HLM 을 통해 조직효과 연구의 대다수는 Level-2 독립변수와 종속변수 (당연히 Level-1 의 Y) 사이의 association 을 다룬다.

2. 이때 Y 는 하나 또는 그 이상의 covariate(s)에 조정된(adjusted) 값이다. 그래서 이러한 연구분석에서 group-mean centering 은 적절하지 않을 것이다.

3. 한가지 주의할 점은 이러한 adjustment 는 가정적 상황에 기초한 것이란 점이다. 다시 말해 전체평균보다 낮은 평균을 가진 집단은 위로, 전체평균보다 높은 집단은 아래로 adjustment 가 이루어지는데 집단간에 평균의 차이가 매우 큰 경우 경험적 현실에 기초하지 않은 순수한 가정일 뿐이라는 점을 염두에 두어야 한다. (예: 여자와 남자의 일 경험을 adjustment 한 후에 일 경험의 회귀계수가 같은 상황에서 이러한 speculation 의 위험: Agresti p.516)

■**Group-mean centered 모형에서는 Level-1 의 "절편"은 위에서 본 바와 같이 unadjusted mean of Y 인데 unadjusted 란 점을 표시하기 위해 μ_j를 사용해서 L-1** 모형을 표시하면

- **Level-1 Model:** $Y_{ij} = \mu_j + \beta_{1j}(X_{ij} - \overline{X}_{\bullet j}) + r_{ij}$

■**Grand-mean centered** 모형은

- **Level-1 Model:** $Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{\bullet\bullet}) + r_{ij}$

$$= \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{\bullet j} + \overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet}) + r_{ij}$$

$$= \beta_{0j} + \beta_{1j}(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet}) + \beta_{1j}(X_{ij} - \overline{X}_{\bullet j}) + r_{ij}$$

두 Level-1 Model 을 비해보면

$$\mu_j = \beta_{0j} + \beta_{1j}(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet}) \rightarrow \beta_{0j} = \mu_j - \beta_{1j}(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})$$

그러므로, Grand-mean centered 모형에서 절편은 각 집단의 평균에서 covariate 의 집단평균과 전체평균의 편차를 반영한 adjustment 를 뺀 값이다.

Model 2. Means as outcomes model with level 1 covariate

위에서 설명한 상황을 반영하기 위해 level-1 covariate 가 있는 means(intercepts)-as-outcomes model 을 보도록 하자. level-1 covariate 는 fixed effects (i.e., assuming that the effects of these covariates are the same for all schools) 일수도 random effects (i.e., assuming that the effects of level 1 variables vary across schools).

LEVEL 1 MODEL

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SECTOR_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

```
Sigma_squared =      36.79508

 Tau
 INTRCPT1,B0      3.96459       0.71641
      SES,B1      0.71641       0.44990

Tau (as correlations)
 INTRCPT1,B0  1.000  0.536
      SES,B1  0.536  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.765
       SES, B1                        0.189
 ----------------------------------------------------

The value of the likelihood function at iteration 21 = -2.330093E+004
 The outcome variable is  MATHACH

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                        Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio    d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.476646   0.231587    49.557       158     0.000
      SECTOR, G01           2.533835   0.344798     7.349       158     0.000
 For      SES slope, B1
    INTRCPT2, G10           2.385451   0.118329    20.160       159     0.000
 ----------------------------------------------------------------------------
```

❑Now the intercept is the value for average SES student in a public school(coded as 0): 11.48. The value for an average-SES Catholic school student is 2.53 units higher: 11.45+2.53=13.98

❑Further, one unit increase in SES is associated with 2.39 units increase in math score.

```
Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect            Standard       Variance     df    Chi-square  P-value
                         Deviation      Component
-----------------------------------------------------------------------------
INTRCPT1,       U0         1.99113        3.96459   158     766.83844    0.000
     SES slope, U1         0.67075        0.44990   159     216.12223    0.002
 level-1,       R          6.06589       36.79508
-----------------------------------------------------------------------------

Statistics for current covariance components model
--------------------------------------------------
Deviance                       = 46601.861400
Number of estimated parameters = 4
```

❑But there is still significant variation across schools in intercepts, and there is also significant variation in SES slopes – so SES doesn’t have the same effect across schools.

Compositional or contextual effects: 조직연구에서 지속적으로 관심을 가지는 것으로 개인수준의 변수를 조직 또는 집단의 수준에서 aggregate 했을 때 가중되는 효과.

Let's take our analysis one step further. Above we considered an individual student's SES as a predictor for Level-1 equation, which was fixed, meaning that it has the same impact on his or her math achievement regardless of the school where that student is studying. Now let's relax that assumption and consider MEANSES as Level-2 predictor.

Model 2. Random-effects ANCOVA with Level-2 predictor

The outcome variable is MATHACH

Summary of the model specified

Level-1 Model

$MATHACH_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(MEANSES_j) + u_{0j}$

$\beta_{1j} = \gamma_{10}$

SES has been centered around the grand mean.

Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*MEANSES_j$
$+ \gamma_{10}*SES_{ij} + u_{0j} + r_{ij}$

Final Results - Iteration 6

Iterations stopped due to small change in likelihood function

$\sigma^2 = 37.01907$

τ

INTRCPT1,β_0 2.69242

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.754

The value of the log-likelihood function at iteration 6 = -2.328429E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.661576	0.149372	84.765	158	<0.001
MEANSES, γ_{01}	3.675037	0.377660	9.731	158	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.191165	0.108667	20.164	7024	<0.001

Hierarchical Linear Models 3

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.661576	0.148394	85.324	158	<0.001
MEANSES, γ_{01}	3.675037	0.352999	10.411	158	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.191165	0.129368	16.938	7024	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	1.64086	2.69242	158	670.17291	<0.001
level-1, r	6.08433	37.01907			

Statistics for current covariance components model

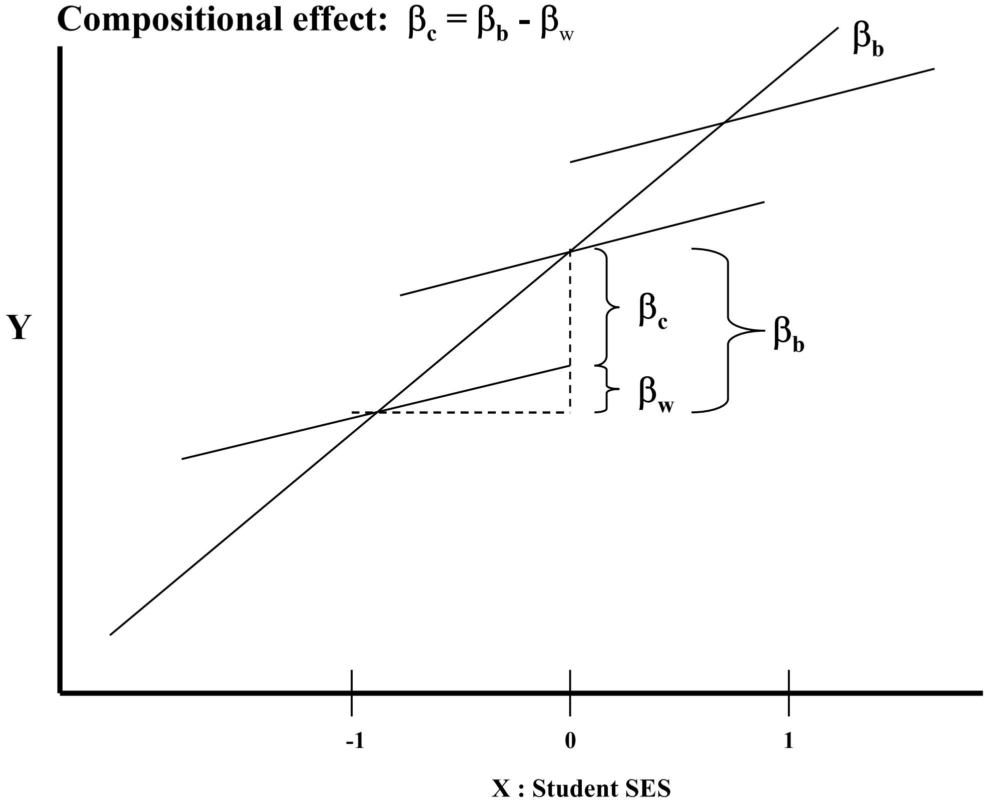
■ Deviance = 46568.579775

Number of estimated parameters = 2

Disentangling Predictors at Two Levels (교재 139~141)

- ❑ **In some cases, variables affect outcomes at both levels—i.e., there is both an individual effect of an predictor and a group-level effect above and beyond the individual level effect**
- ❑ **The latter is known as a compositional(contextual) effect**
  - **--E.g., SES**
  - **--Why might there be a compositional effect of SES?**
    1. **Peer effects**
    2. **Proxy for other variables not in model**

Compositional effect: $\beta_c = \beta_b - \beta_w$



β_w **= expected difference in Y between two students in same school who differ by one unit in SES**

β_b **= expected difference in mean of Y(즉, Y-bar) between two schools that differ by one unit of mean SES**

β_c **= difference in expected Y between two students who have the same individual SES, but who attend schools that differ by one unit of mean SES**

❑ **정리하면 "**β_c**"는 SES 가 한 단위 아래인 학교에 비해 SES 가 한 단위 위인 학교에 다니는 학생이 얻게 되는 수학점수의 증가 량**

Two ways to estimate (센터링을 달리한 회귀계수 값을 보자!)

1. **Grand mean centering**

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{\bullet\bullet}) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\overline{X}_{\bullet j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$\gamma_{01} = \beta_c$ = **compositional effects**

$\gamma_{10} = \beta_w$ = **person-level effects**

$\beta_b = \gamma_{01} + \gamma_{10}$ = **school-level (aggregate school level regression) effects**

2. **Group mean centering**

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{\bullet j}) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\overline{X}_{\bullet j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$\gamma_{01} = \beta_b$ = **school-level effects (데이터를 aggregate 했을 때 회귀계수)**

$\gamma_{10} = \beta_w$ = **person-level effects (within school 에서 회귀계수)**

$\beta_c = \gamma_{01} - \gamma_{10}$ = **compositional effects**

요약하면 분석에서 Group-mean centering 을 선택하면 X_{ij} 와 Y_{ij} 사이의 관계는 집단 내와 집단 간 요소로 구분돼서 나타난다. 그래서 compositional effect 를 구하려면 위에서처럼 빼기 연산을 해야 한다. ($\beta_c = \gamma_{01} - \gamma_{10}$)

The outcome variable is MATHACH

Summary of the model specified

Level-1 Model

$MATHACH_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(MEANSES_j) + u_{0j}$

$\beta_{1j} = \gamma_{10}$

SES has been centered around the group mean.

Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*MEANSES_j$
$+ \gamma_{10}*SES_{ij} + u_{0j} + r_{ij}$

Final Results - Iteration 6

Iterations stopped due to small change in likelihood function

$\sigma^2 = 37.01907$

τ

INTRCPT1,β_0 2.69235

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.754

The value of the log-likelihood function at iteration 6 = -2.328429E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.648117	0.149370	84.677	158	<0.001
MEANSES, γ_{01}	5.866175	0.361686	16.219	158	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.191172	0.108667	20.164	7024	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	$d.f.$	χ^2	p-value
INTRCPT1, u_0	1.64084	2.69235	158	670.16369	<0.001
level-1, r	6.08433	37.01907			

Statistics for current covariance components model

Deviance = 46568.577332

Number of estimated parameters = 2