Description of the Example Dataset

- High School and Beyond Survey (**hsb**),
- Consists of 7185 students from 160 schools (**sch.id**),
- Sample sizes average about 45 students per school,
- Math achievement as DV: **math** (Level 1),
- IV (Level 1):
  - **minority** (1 = minority, 0 = others)
  - **female** (1 = female, 0 = male)
  - **ses** (numeric)
- IV (Level 2):
  - **size** (school size)
  - **meanses** (school ses)
  - **sector** (1 = Catholic, 0 = Public)

	sch.id	math	size	sector	meanses	minority	female	ses
1	1224	5.876	842	0	-0.428	0	1	-1.528
2	1224	19.708	842	0	-0.428	0	1	-0.588
3	1224	20.349	842	0	-0.428	0	0	-0.528
4	1224	8.781	842	0	-0.428	0	0	-0.668
5	1224	17.898	842	0	-0.428	0	0	-0.158
6	1224	4.583	842	0	-0.428	0	0	0.022

Comparing HLM M-o-M model with non-HLMs

- Let's look at a binary (categorical), school-level predictor, `sector`:

```
• hsb <- read_csv('https://raw.githubusercontent.com/rnorouzian/e/master/hsb.csv')
• hsb <- mutate(hsb, sector = factor(ifelse(sector==0, "pub", "cath")))
• g <- hsb %>% dplyr::select(math, sector) %>% group_by(sector) %>%
  summarise(across(.fns = list(mean=mean, sd=sd)), n = n())
```

sector	math_mean	math_sd	n
cath	14.17030	6.359018	3543
pub	11.36407	7.079920	3642

```
• t.testb(m1=g$math_mean[1], m2=g$math_mean[2], s1=g$math_sd[1], s2=g$math_sd[2], n1=
  g$n[1], n2=g$n[2], var.equal = TRUE)
```

mean.dif	std.error	t.value	p.value
-2.806225	0.158905	-17.6598	0

- You will get exactly the same SE if you do: `summ(lm(math~ sector, data = hsb))` # WHY so?

Unbiased SE (Estimated by HLM)

- Level 1: $MathAch_{ij} = \beta_{0j} + e_{ij}$
- Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} Sector_j + U_{0j}$

Combined:

- $MathAch_{ij} = \gamma_{00} + \gamma_{01} Sector_j + U_{0j} + e_{ij}$

```
• hsb <- mutate(hsb, sector= relevel(sector, ref= "pub"))
• summ(lmer(math~ sector+ (1|sch.id), data= hsb),re.variance= "var")
```

```
------------------------------------------------------------------
                          Est.       S.E.      t val.        d.f.          p
------------------- ---------- --------- ---------- ----------- ----------
(Intercept)           11.39304    0.29283   38.90657   158.53713    0.00000
sectorcath             2.80489    0.43906    6.38845   153.50058    0.00000
------------------------------------------------------------------
```

Coefficient's (γ_{00}) interpretation: The mean of the "public" high schools.
Coefficient's (γ_{01}) interpretation: Mean_cath – Mean_pub (mean diff.).
So, the mean of **Cath_HS**: **11.39304 + 2.80489 = 14.19793**.

Sampling variance and SE of coef.s in HLM

```
RANDOM EFFECTS:
--------------------------------------
  Group       Parameter        Var.
---------- ------------- ----------
  sch.id     (Intercept)     6.67696
 Residual                   39.15140
--------------------------------------
```

$$\mathrm{VAR}(\hat{\gamma}_{01}) = \frac{\tau_{00} + \frac{\sigma_e^2}{n}}{J\sigma_{sector}^2}$$

$\mathrm{VAR}(U_{0j}) = \tau_{00} = 6.68$

$\mathrm{VAR}(e_{ij}) = \sigma_e^2 = 39.15$

$\mathrm{VAR}(\text{sector}) = \sigma^2_{\text{sector}} \approx .25$ → `var(as.numeric(hsb$sector))`

n = Number of students in each school (average per school) ≈ 45

```
hsb %>% dplyr::select(sch.id) %>% group_by(sch.id) %>% summarise(n=n())
%>% summarise(n=mean(n))
```

J = Total number of schools = 160

$$SE_{\hat{\gamma}_{01}} = \sqrt{\frac{\tau_{00} + \frac{\sigma_e^2}{n}}{J\sigma_{sector}^2}} = \sqrt{\frac{6.68 + \frac{39.15}{45}}{160(.25)}} = \mathbf{.439}$$

(Correct)

Sampling variance and SE in non-HLM

- Without considering the multilevel/grouping structure (***Exception Fallacy***):
- Earlier, we noted if τ_{00} is ignored, where does it end up going?
- $*\sigma^2_e$ (*residual variance*): `sigma(lm(math~sector,data = hsb))^2 #`= 45.35

- $*\sigma^2_e$ is approximated by the sum of the following:
- $\mathrm{VAR}(U_{0j}) = \tau_{00} = 6.68$
- $\mathrm{VAR}(e_{ij}) = \sigma_e^2 = 39.15$

$$SE_{\hat{\gamma}_{01}} = \sqrt{\frac{\dfrac{\underline{*\sigma_e^2}}{n}}{J\sigma^2_{sector}}} \approx \sqrt{\frac{\dfrac{\underline{(\tau_{00} + \sigma^2{}_e)}}{n}}{J\sigma^2_{sector}}} = \sqrt{\frac{\dfrac{\underline{45.35}}{45}}{160(.25)}} = \mathbf{.1587}$$

(Incorrect)

HLM and its benefits

- MLM can take the non-independency (between lower level observations) into account:

- Estimate the cluster effect,
- Obtain the correct standard errors for the parameter estimates,
- Have correct statistical tests for the parameter estimates.

- The *main purpose* of using MLM is NOT that MLM has more statistical power! On the other hand, you can increase the statistical power in MLM by…

- Increasing the # of schools (J) better reduces SE of a coef. than the # of students (n). Why?

$$VAR(\hat{\gamma}_{01}) = \frac{\tau_{00} + \frac{\sigma_e^2}{n}}{J\sigma^2_{sector}} = \left(\frac{\tau_{00}}{J} + \frac{\sigma_e^2}{nJ}\right)\frac{1}{\sigma^2_{sector}}$$

- Because of their positions in the formula, J reduces the entire formula but n reduces sigma squared.

- Reducing the size of VAR(γ_{01}) results in the INCREASE in POWER of detecting significant γ_{01}.

Side by Side Output Comparisons

- `m1 <- lmer(math~ sector + (1|sch.id), data= hsb)`
- `ols1 <- lm(math~ sector, data = hsb)`
- `stargazer(ols1, m1, type = 'text', star.cutoffs = c(.05,.01,.001))`

```
=============================================================
                              Dependent variable:
                    -----------------------------------------
                                     math
                              OLS              linear
                                            mixed-effects
                              (1)               (2)
-------------------------------------------------------------
sector                      2.806***          2.805***
                            (0.159)           (0.439)
Constant                    11.364***         11.393***
                            (0.112)           (0.293)
-------------------------------------------------------------
Observations                 7,185             7,185
R2                           0.042
Adjusted R2                  0.041
Log Likelihood                              -23,540.070
Akaike Inf. Crit.                            47,088.130
Bayesian Inf. Crit.                          47,115.650
Residual Std. Error     6.734 (df = 7183)
F Statistic         311.869*** (df = 1; 7183)
=============================================================
Note:                           *p<0.05; **p<0.01; ***p<0.001
```

Means-as-outcomes model (*categorical predictor*)

- Level 1: $MathAch_{ij} = \beta_{0j} + e_{ij}$
- Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} Sector_j + U_{0j}$

Combined:

- $MathAch_{ij} = \gamma_{00} + \gamma_{01} Sector_j + U_{0j} + e_{ij}$

```
hsb <- mutate(hsb, sector= relevel(sector, ref= "pub"))
summ(lmer(math~ sector+ (1|sch.id), data= hsb),re.variance= "var")
```

	Est.	S.E.	t val.	d.f.	p
(Intercept)	11.39304	0.29283	38.90657	158.53713	0.00000
sectorcath	2.80489	0.43906	6.38845	153.50058	0.00000

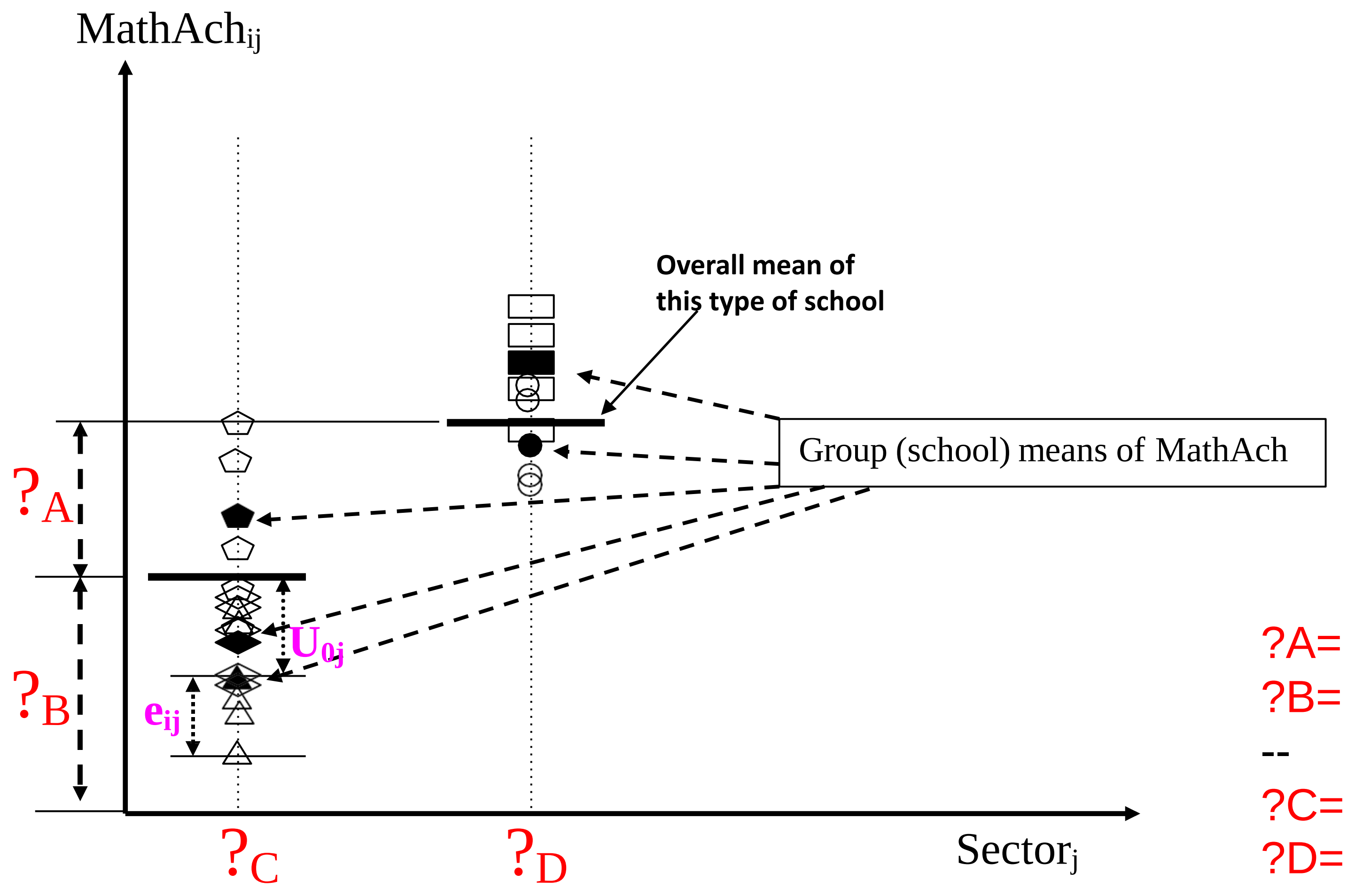
RANDOM EFFECTS:

Group	Parameter	Var.
sch.id	(Intercept)	6.68
Residual		39.15

$^*\tau_{00}$ = 8.61 → random-intercept model's bet. School variance
$^*\sigma^2$ = 39.14 → random-intercept model's within variance

($^*\tau_{00}$ - τ_{00}) / $^*\tau_{00}$ = (8.61 – 6.68) / 8.61 = .225 (or 22.5%) decrease in between-school variance compared to Model 1 as a result of adding sector as a predictor of bet. school means' variation!

Visualizing Means-as-outcomes model *(categorical predictor)*



- Level 1: $MathAch_{ij} = \beta_{0j} + e_{ij}$
- Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}Sector_j + U_{0j}$

Combined:

- $MathAch_{ij} = \gamma_{00} + \gamma_{01}Sector_j + U_{0j} + e_{ij}$

?A= Difference bet. “pub” & “cath” overall means (γ_{01})
?B= Reference schools’ overall mean (γ_{00})
--
?C= 1st type of schools; "Public" (coded “pub”)
?D= 2nd type of schools; "Catholic" (coded “cath”)

Random-Coefficients Regression Model

- **RQ3:** On average, does student SES relate to math achievement? Is this relation similar across schools?

- Pretend that we regress **math** on **ses** for each school; in other words, we would run 160 regressions, each for a school.

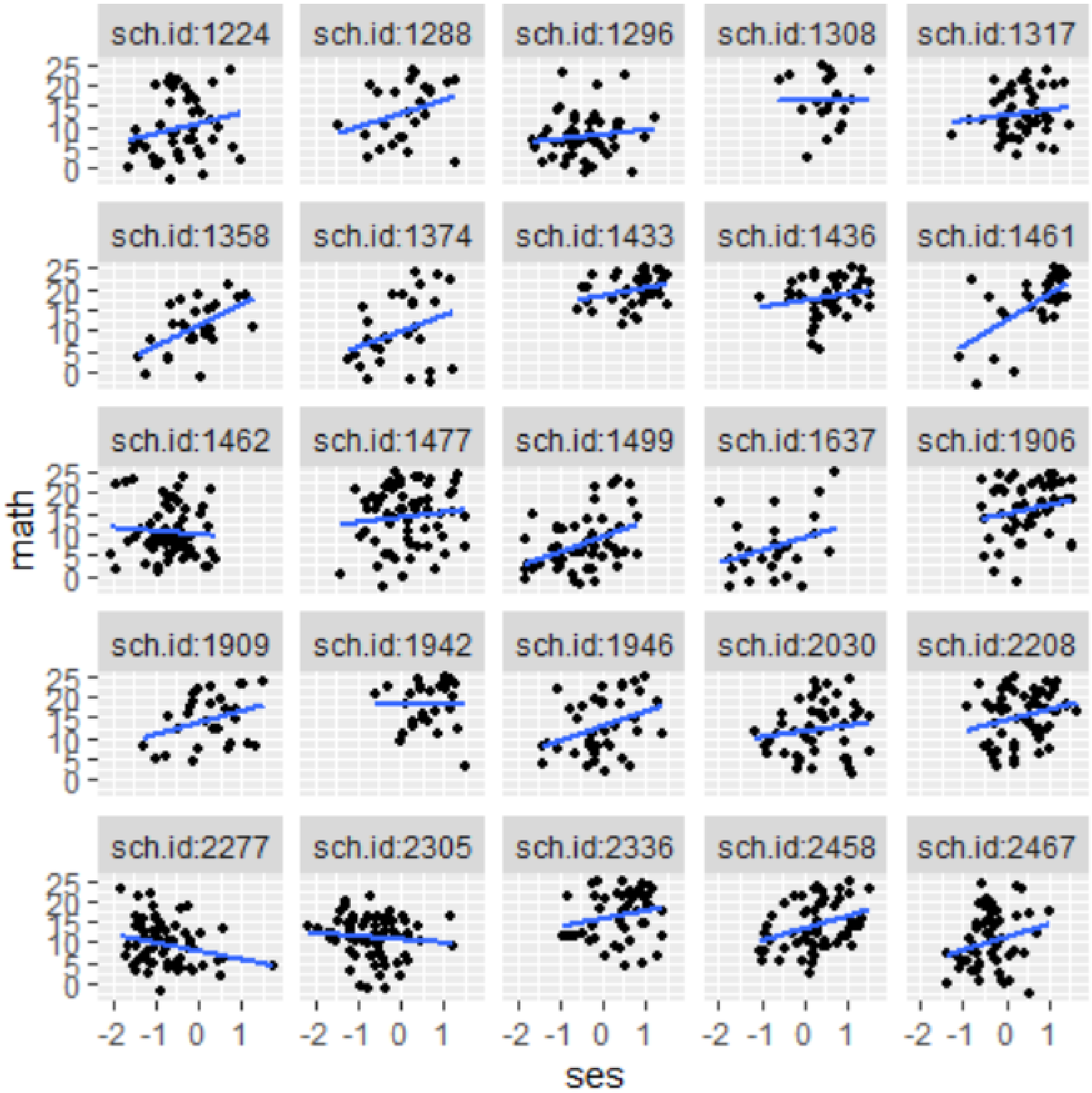
- What would be the average of the 160 regression equations (*both* intercepts and slopes)?

- How much do the regression equations vary from school to school?

- What is the correlation between the *intercepts* and *slopes* across the schools?

Does student SES relate to math achievement? Is this relation similar across schools?

```
• some <- subset(hsb, sch.id %in%
  unique(sch.id)[1:25])

• ggplot(some) + aes(ses, math)+
  geom_point() + facet_wrap(~sch.id)
  geom_smooth(method="lm",se=F)
```



Random-Coefficients Regression Model

sub$_0$ = intercept; sub$_1$ = slope

Level 1: $\text{MathAch}_{ij} = \beta_{0j} + \beta_{1j}\,\text{SES}_{ij} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + U_{0j}$ → Intercept deviations from overall mean

$\beta_{1j} = \gamma_{10} + U_{1j}$ → Slope deviations from overall model slope

Combined:

$$\text{MathAch}_{ij} = \gamma_{00} + \gamma_{10}\,\text{SES}_{ij} + U_{0j} + U_{1j}\,\text{SES}_{ij} + e_{ij}$$

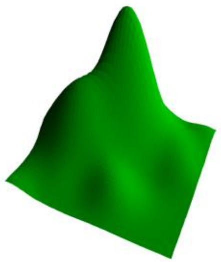
$$G = \text{Cov}(U_{0j}, U_{1j}) = \begin{array}{c c} & \begin{array}{cc} U_{0j} & U_{1j} \end{array} \\ \begin{array}{c} U_{0j} \\ U_{1j} \end{array} & \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \end{array} \rightarrow$$

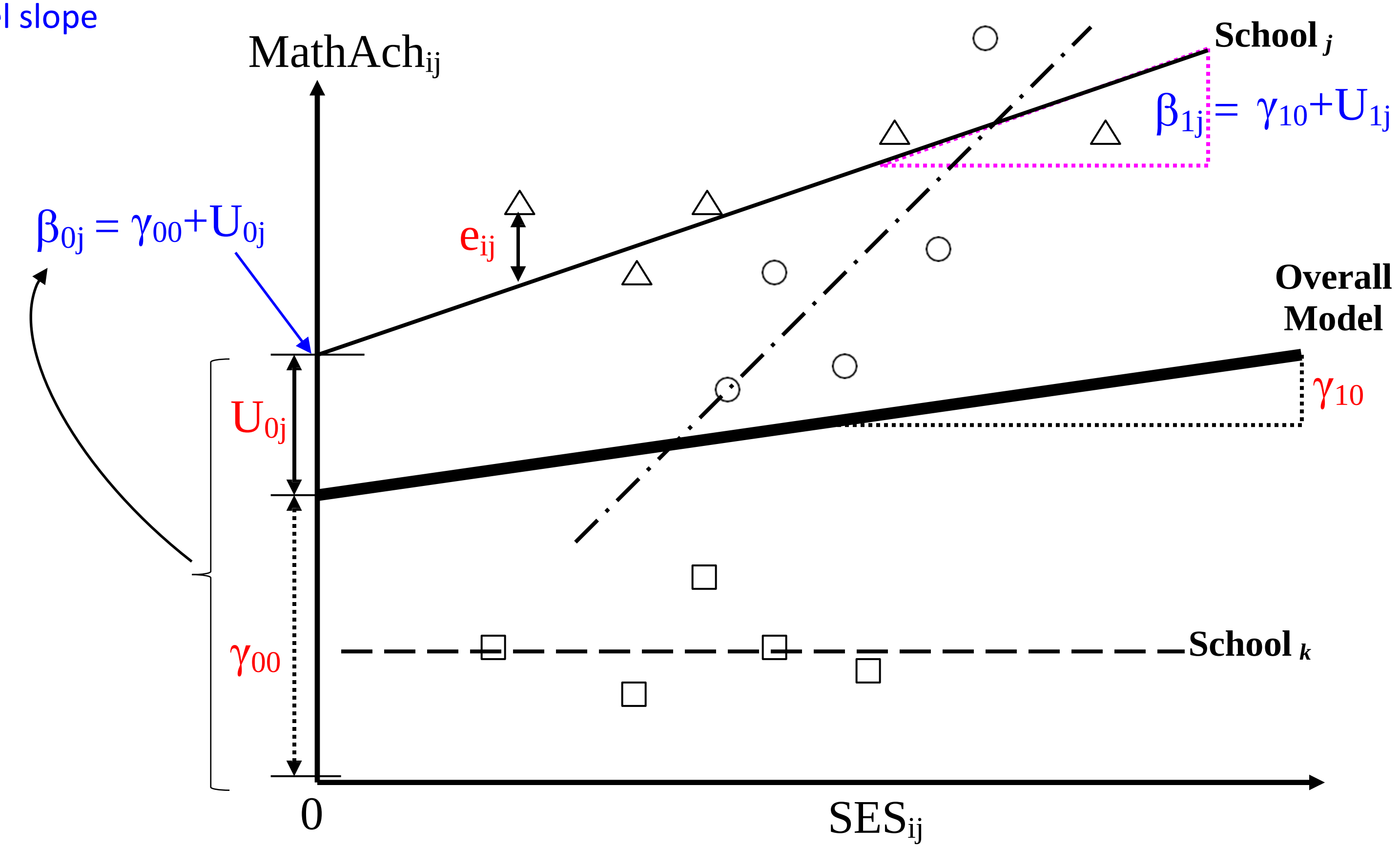
Diagonal = variances

Off-diag. = covariance

$\text{Cov}(e_{ij}, U_j) = 0$

- Bet. variation(s) and within variation are separate and additive.
- $U_j \sim \text{mv}Norm(\mathbf{0}, G)$ →





Correlation bet. U_js = Correlation bet. β_js

- *G* matrix represents the correlation bet. β_js (slopes & intercepts) and not just that bet. U_js (their deviations from the β_js), WHY so?
- Additive constants (γ_{00} & γ_{10}) don't change the correlations.

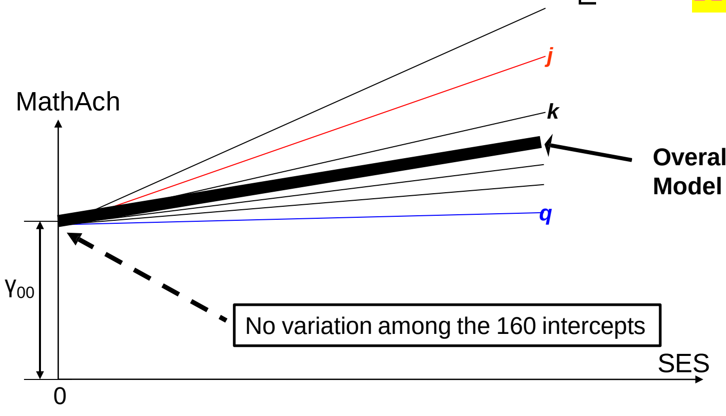
```
# Additive constants
gamma_00 = 10
gamma_10 = 2

G = matrix(c(4,.2,.2,5),2)  #e.g.,
```

$\rightarrow G = \begin{bmatrix} 4 & .2 \\ .2 & 5 \end{bmatrix}$

```
set.seed(1)
Uj = mvrnorm(9, mu = c(0,0), Sigma = G)

cor(Uj) # .205

B0j = gamma_00 + Uj[,1]
B1j = gamma_10 + Uj[,2]

cor(B0j,B1j) # .205
```

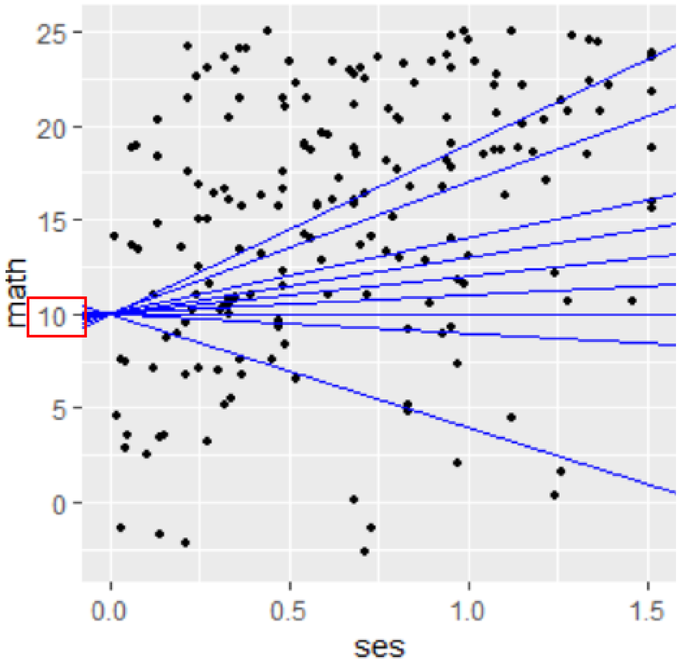
Meaning of τ_{00}, τ_{11}, and $\tau_{01}(=\tau_{10})$

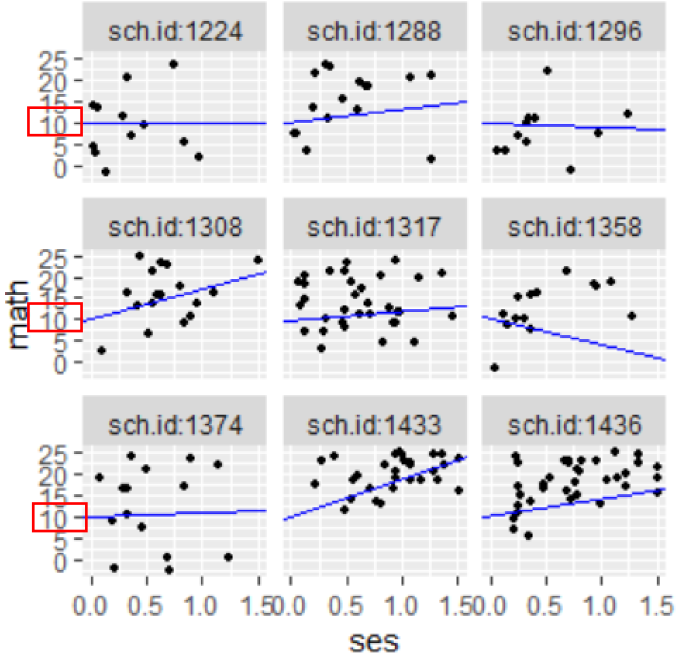
$$G = \begin{array}{c|cc} & U_{0j} & U_{1j} \\ \hline U_{0j} & 0 & 0 \\ U_{1j} & 0 & \tau_{11} \end{array}$$



- Variation in slopes (τ_{11}) across schools but not intercepts (i.e., a constant).
  Thus, **Cov(slopes, constant) = 0.**

```
hsb <- read_csv('https://raw.githubusercontent.com/rnorouzian/e/master/hsb.csv')
nine <- subset(hsb, sch.id %in% unique(sch.id)[1:9])  # get 9 schools for display

ggplot(nine) + aes(ses, math)+ geom_point() + xlim(0,NA)+
  geom_abline(intercept=10,slope = c(0,3,-1,7,2,-6,1,9,4), color="blue")
```

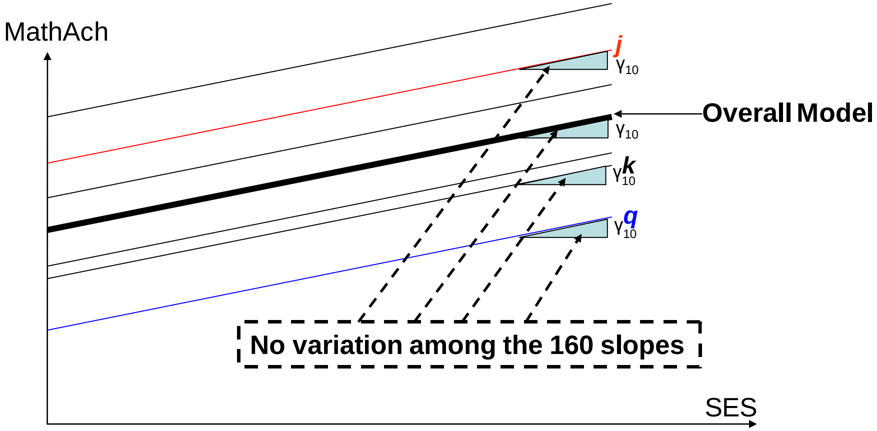




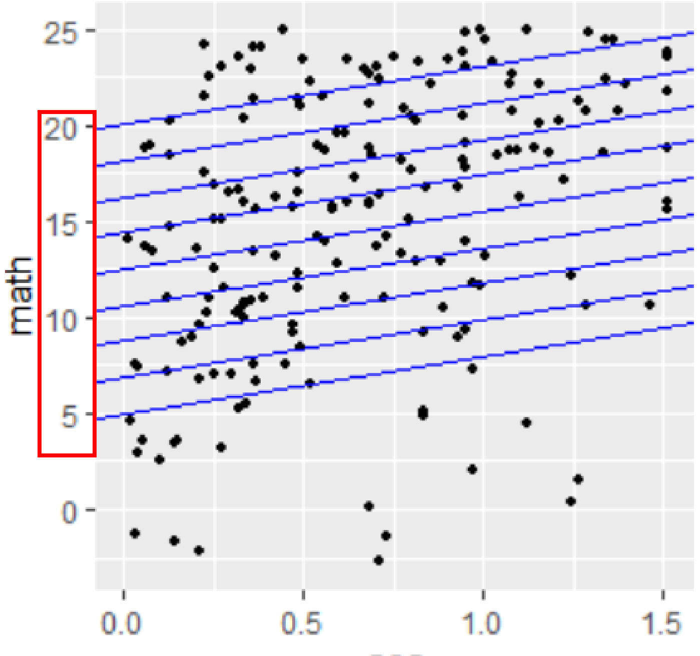
Meaning of τ_{00}, τ_{11}, and $\tau_{01} (= \tau_{10})$

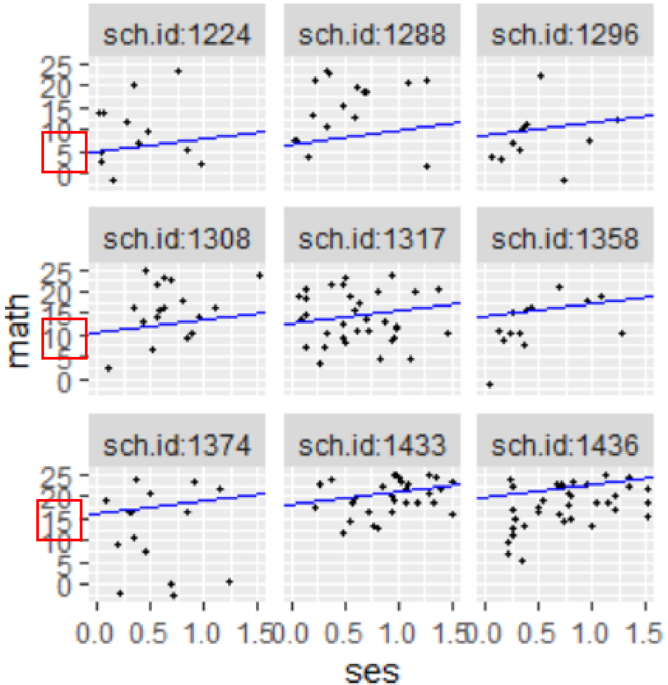
$$G = \begin{array}{c|cc} & U_{0j} & U_{1j} \\ \hline U_{0j} & \tau_{00} & 0 \\ U_{1j} & 0 & 0 \end{array}$$

- Variation in intercepts (τ_{00}) across schools but not slopes (i.e., a constant). Thus, **Cov(intercepts, constant) = 0.**



- `ggplot(nine) + aes(ses, math)+ geom_point() + xlim(0,NA)+ geom_abline(intercept=seq(5,20,length= 9),slope= 3, color="blue")`

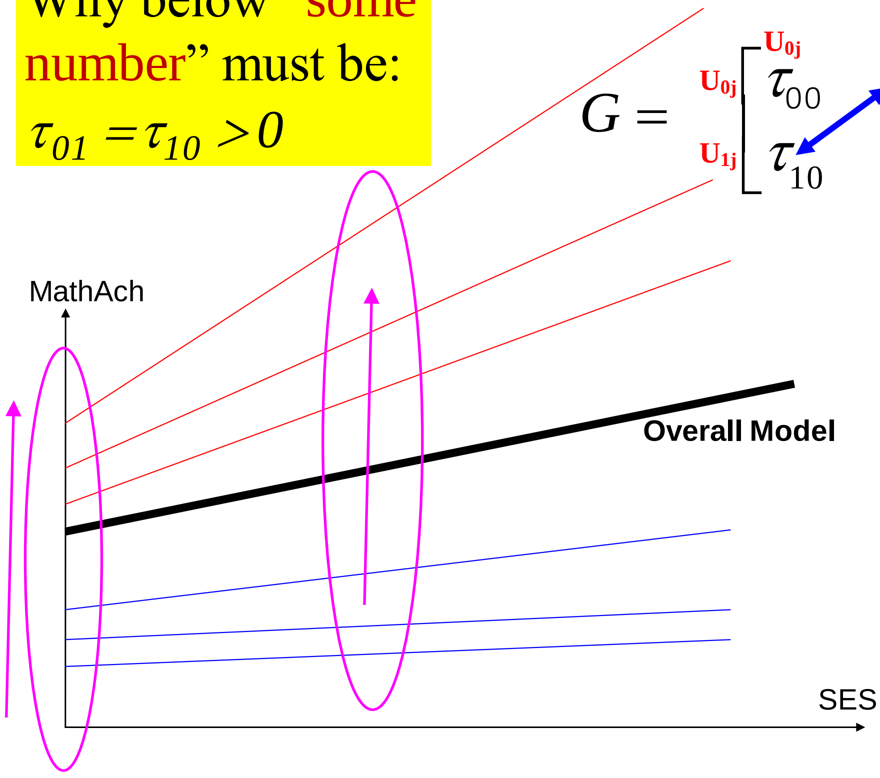




Meaning of τ_{00}, τ_{11}, and $\tau_{01}(=\tau_{10})$

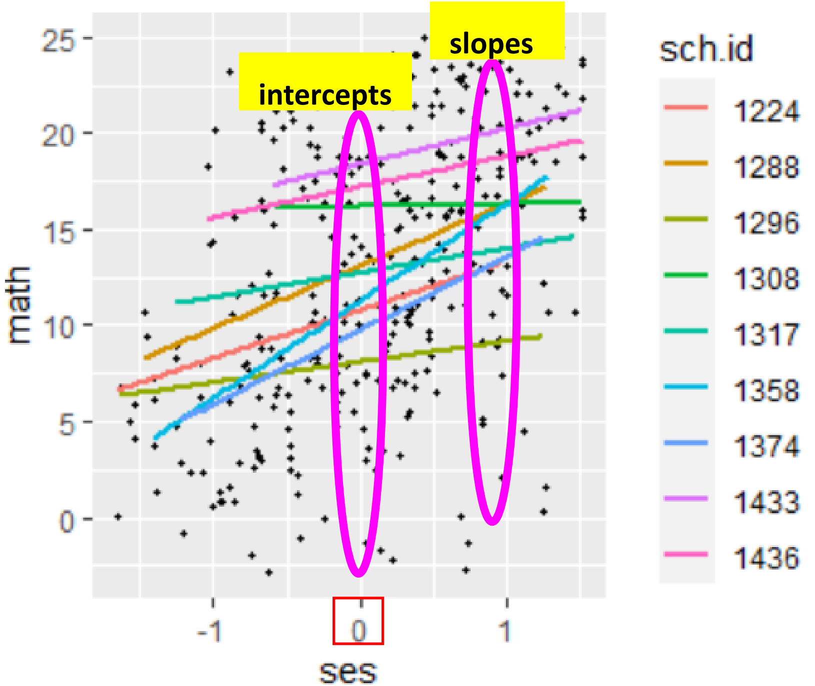
Why below "some number" must be:
$\tau_{01} = \tau_{10} > 0$

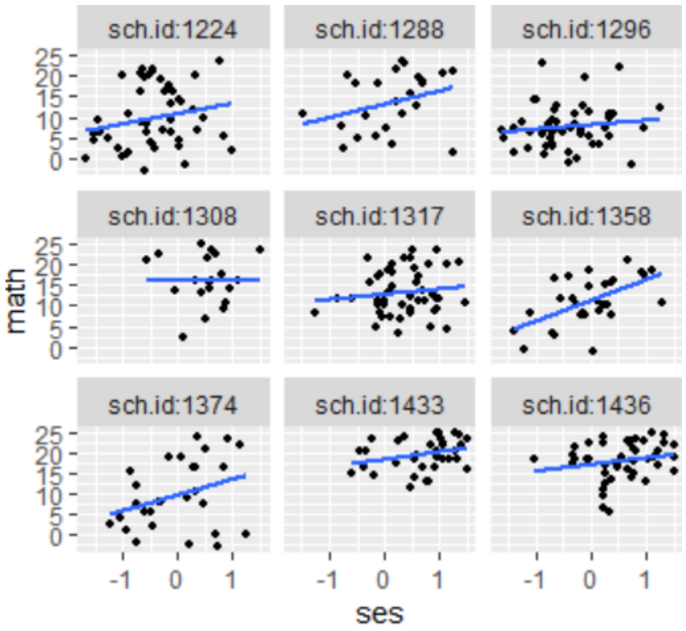
$$G = \begin{array}{c} \\ U_{0j} \\ U_{1j} \end{array} \begin{array}{cc} U_{0j} & U_{1j} \\ \left[\begin{array}{cc} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{array}\right] \end{array}$$



- Variation both in intercepts and slopes. AND,
  Cov(intercepts, slopes) = some number.

```
• nine$sch.id <- factor(nine$sch.id)
• ggplot(nine) + aes(ses, math)+ geom_point() +
    geom_smooth(method="lm",se=F, aes(color=sch.id))
```



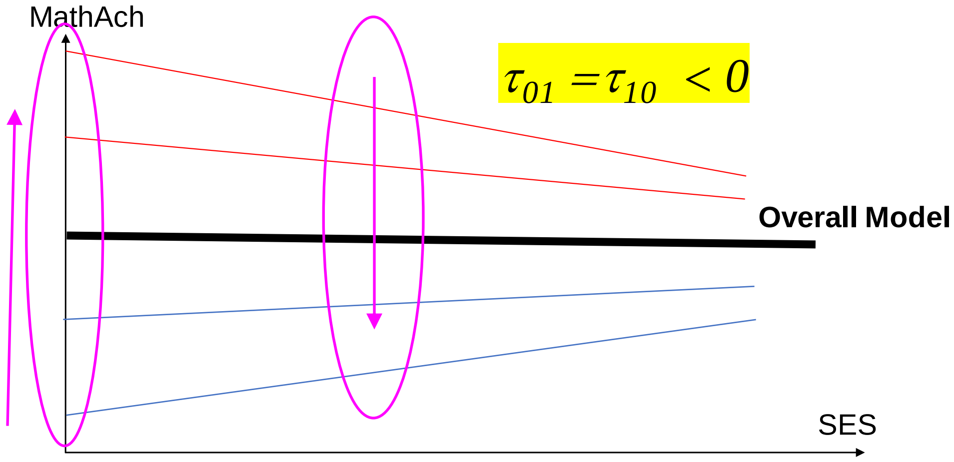


- In schools with higher mean mathach (intercept), **ses** has a higher effect (slope) on math achievement.

A Question for You?

- What do the following visuals tell us about the type of relationship between the intercepts & slopes of schools?

$$G = \begin{array}{c c} & \begin{array}{cc} U_{0j} & U_{1j} \end{array} \\ \begin{array}{c} U_{0j} \\ U_{1j} \end{array} & \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \end{array}$$



- In schools with higher mean mathach (intercept), ses has a lower effect (slope) on math achievement.