

# Хијерархијски линеарни модел

- \* Кристина Матовић
- \* Марија Костић
- \* Ана Антић

# Хијерархијски модели у образовању



Слика 1: Пример хијерархијске структуре у образовању

# Проблеми

- \* Над статистикама везаним за образовање, где подаци имају природну хијерархијску структуру, јављајају се одређени проблеми приликом употребе обичне линеарне регресије.
- \* Проблем игнорисања важности ефеката групе и претпоставка о независности.
- \* Проблем са „cross-level “ ефектима

# Хијерархијски линеарни модели на два нивоа

- \* Описиваћемо проблеме са хијерархијом на два нивоа, иако је, такође, могуће вршити хијерархију на више нивоа.
- \* Имамо податке за  $j = 1, \dots, M$  група (школа) и различит број  $n_j$  појединаца (ученика) у свакој од тих групи.
- \* Подаци не морају нужно бити уравнотежени.
- \* На нивоу ученика (ниво један) имамо зависну променљиву  $Y_{ij}$  и променљиву објашњења (то јест независну променљиву)  $x_{ij}$ .
- \* На нивоу школе (ниво два) такође имамо променљиву објашњења  $W_j$ .

# Ниво 1 регресионе једначине

$$* \quad Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$$

- \*  $Y_{ij}$  представља исход за  $i$  – тог ученика у  $j$  –тој школи тј. зависна променљива за појединачно посматрање на нивоу 1
- \*  $x_{ij}$  односи се на предиктор нивоа 1 тј. објашњава променљиву за  $i$  – тог ученика у  $j$  –тој школи
- \*  $\beta_{0j}$  је коефицијент регресије који се односи на пресретање зависне променљиве у  $j$  –тој групи тј. коефицијент пресретање за  $j$  –ту школу
- \*  $\beta_{1j}$  је коефицијент регресије који се односи на нагиб за  $j$  –ту школу
- \*  $\varepsilon_{ij}$  је случајна грешка предвиђања за  $i$  – тог ученика у  $j$  –тој школи

# Ниво 2 регресионе једначине

- \*  $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$
- \*  $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$
- \*  $\gamma_{00}$  просечно пресретање када је  $W_j = 0$  тј. то је средња вредност зависне променљиве у свим групама када су сви предиктори једнаки нули
- \*  $\gamma_{01}$  укупни коефицијент регресије, или нагиба, између зависне променљиве и предиктора нивоа 2
- \*  $u_{0j}$  случајна грешка (одступање коефицијента пресретања од просечног пресретања) за  $j$  —ту школу
- \*  $\gamma_{10}$  укупни коефицијент регресије, или нагиба, између зависне променљиве и предиктора нивоа 1 када је  $W_j = 0$
- \*  $u_{1j}$  случајна грешка (одступање коефицијента нагиба од просечног нагиба) за  $j$  —ту школу

# Једначина модела

- \*  $Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}W_j + \gamma_{11}W_jx_{ij} + u_{0j} + u_{1j}x_{ij} + \varepsilon_{ij}$
- \* Ефекат интеракције на више нивоа (*cross – level interaction effect*):  $\gamma_{11}W_jx_{ij}$
- \* Фиксни део модела:  $\gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}W_j + \gamma_{11}W_jx_{ij}$
- \* Случајни део модела:  $u_{0j} + u_{1j}x_{ij} + \varepsilon_{ij}$

# Претпоставке модела

- \* Линеарност
- \* Нормираност
- \* Хомоскедастичност  $D(\varepsilon_{ij}) = \sigma^2 > 0$
- \*  $x_{ij}$  и  $\varepsilon_{ij}$  су независни
- \*  $E(\varepsilon_{ij}) = 0$  ,  $D(\varepsilon_{ij}) = \sigma^2$
- \*  $E(\beta_{0j}) = \gamma_{00}$  ,  $var(\beta_{0j}) = \tau_{00} = \tau_0^2$
- \*  $E(\beta_{1j}) = \gamma_{10}$  ,  $var(\beta_{1j}) = \tau_{11} = \tau_1^2$
- \* Грешке нивоа 1 и нивоа 2 су хомогене и некорелисане тј.  
 $E(\varepsilon_i \varepsilon_j) = 0$  за  $i \neq j$



# Испитивање података на основу базе

- \* Базе MathAchieve и MathAchSchool.

```
head(MathAchieve,10)
```

```
## Grouped Data: MathAch ~ SES | School
```

##	School	Minority	Sex	SES	MathAch	MEANSES
## 1	1224	No	Female	-1.528	5.876	-0.428
## 2	1224	No	Female	-0.588	19.708	-0.428
## 3	1224	No	Male	-0.528	20.349	-0.428
## 4	1224	No	Male	-0.668	8.781	-0.428
## 5	1224	No	Male	-0.158	17.898	-0.428
## 6	1224	No	Male	0.022	4.583	-0.428
## 7	1224	No	Female	-0.618	-2.832	-0.428
## 8	1224	No	Male	-0.998	0.523	-0.428
## 9	1224	No	Female	-0.888	1.527	-0.428
## 10	1224	No	Male	-0.458	21.521	-0.428

```
head(MathAchSchool,10)
```

##	School	Size	Sector	PRACAD	DISCLIM	HIMINTY	MEANSES
## 1224	1224	842	Public	0.35	1.597	0	-0.428
## 1288	1288	1855	Public	0.27	0.174	0	0.128
## 1296	1296	1719	Public	0.32	-0.137	1	-0.420
## 1308	1308	716	Catholic	0.96	-0.622	0	0.534
## 1317	1317	455	Catholic	0.95	-1.694	1	0.351
## 1358	1358	1430	Public	0.25	1.535	0	-0.014
## 1374	1374	2400	Public	0.50	2.016	0	-0.007
## 1433	1433	899	Catholic	0.96	-0.321	0	0.718
## 1436	1436	185	Catholic	1.00	-1.141	0	0.569
## 1461	1461	1672	Public	0.78	2.096	0	0.683

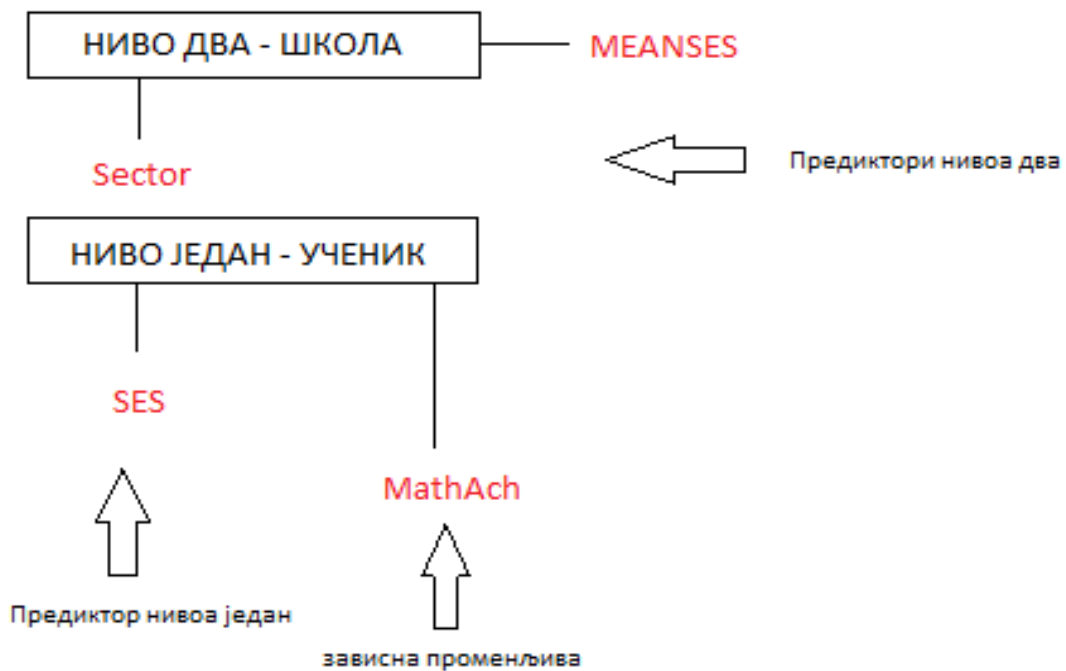
# Коришћене променљиве

- \* School
- \* SES
- \* MathAch
- \* Meanses
- \* Cses

# Нова база

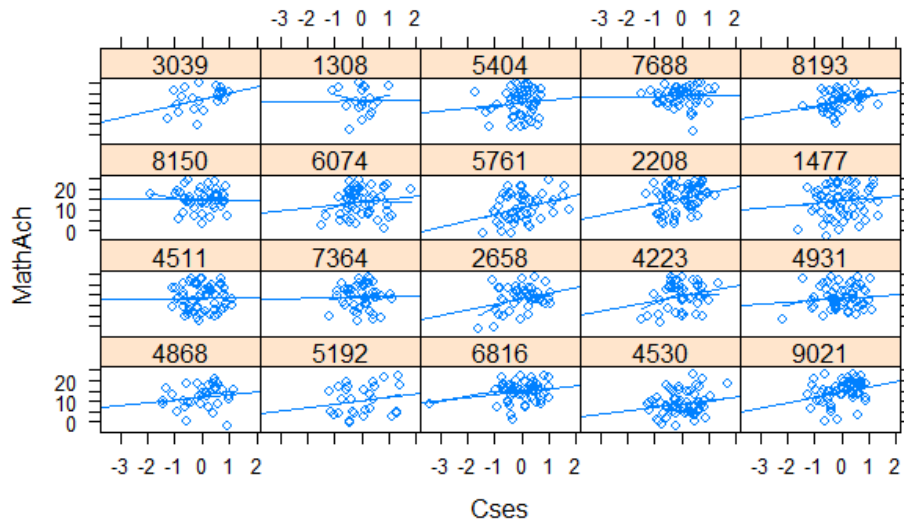
```
Math[sample1, ]
```

##	School	SES	MathAch	Sector	Cses	Meanses
## 895	2305	0.182	12.313	Catholic	0.8100000	-0.6280000
## 1208	2629	0.502	15.180	Catholic	0.6396491	-0.1376491
## 2115	3427	0.672	20.621	Catholic	0.5189796	0.1530204
## 2354	3657	1.512	18.610	Public	2.1611765	-0.6491765
## 2766	4042	0.772	22.923	Catholic	0.3700000	0.4020000
## 4146	5838	-0.038	13.903	Public	-0.1945161	0.1565161
## 4324	6144	-0.808	13.247	Public	-0.3704651	-0.4375349
## 4340	6170	1.002	19.951	Public	1.3038095	-0.3018095
## 4528	6415	0.942	13.091	Public	1.1292593	-0.1872593
## 5802	8150	0.902	16.405	Catholic	0.5961364	0.3058636

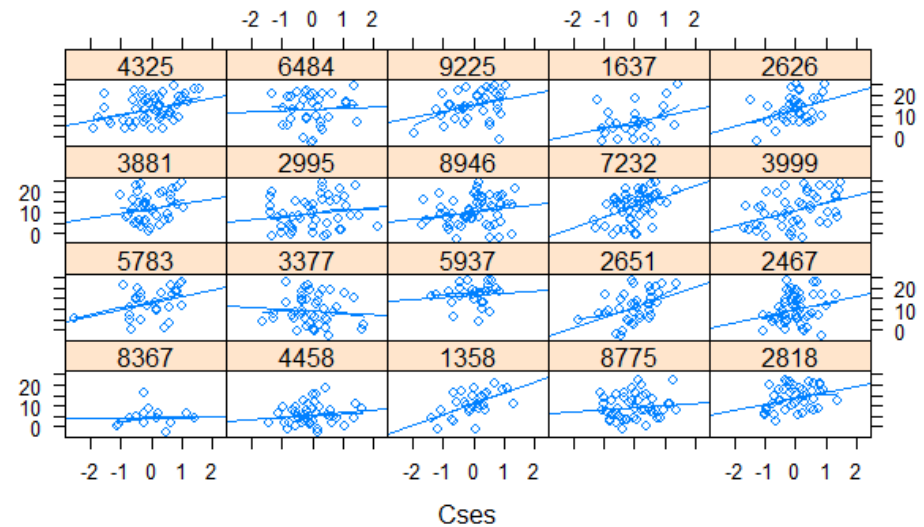


# Католичке и јавне школе

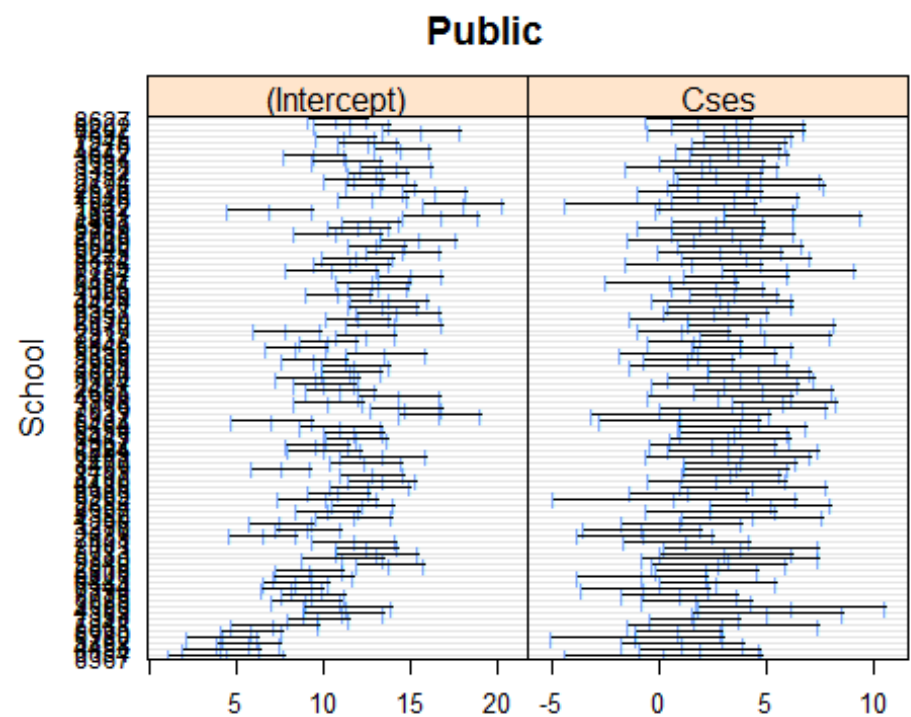
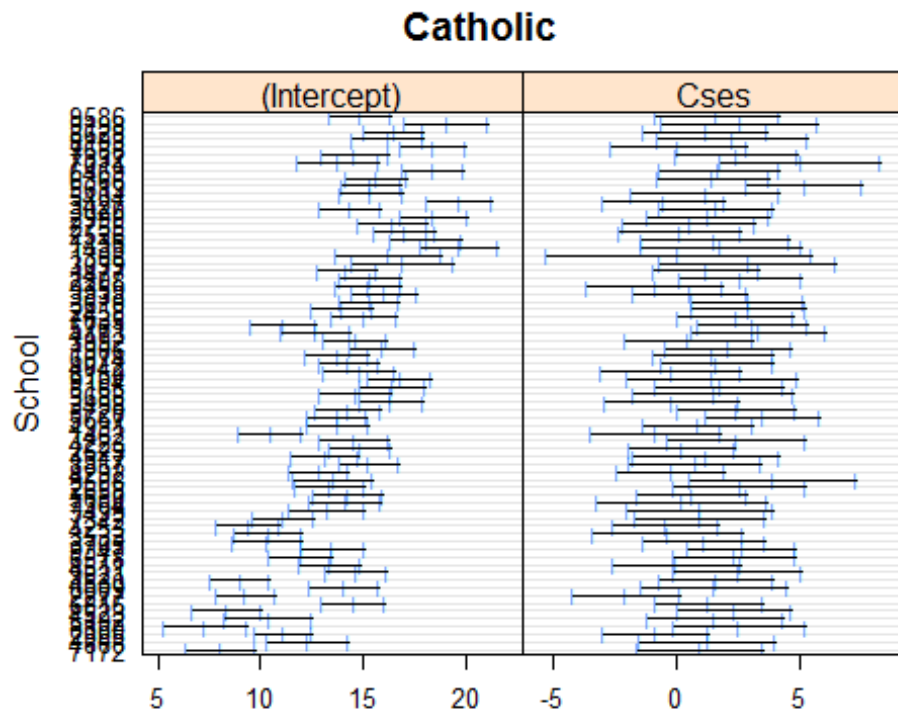
Catholic



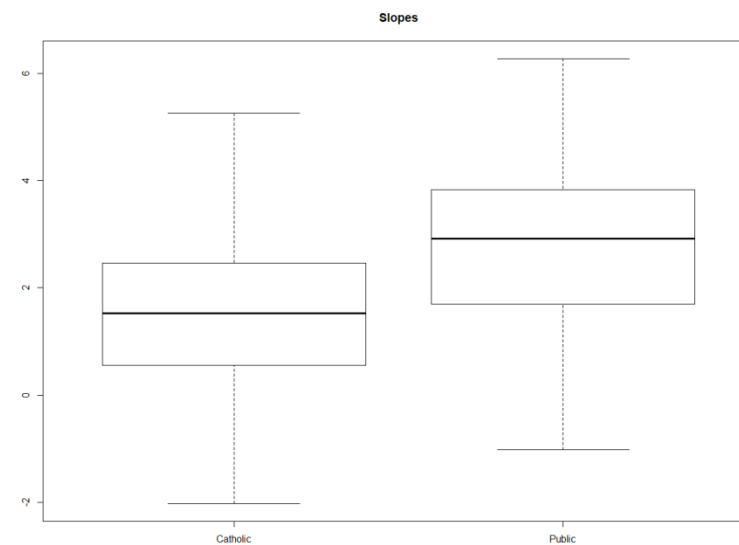
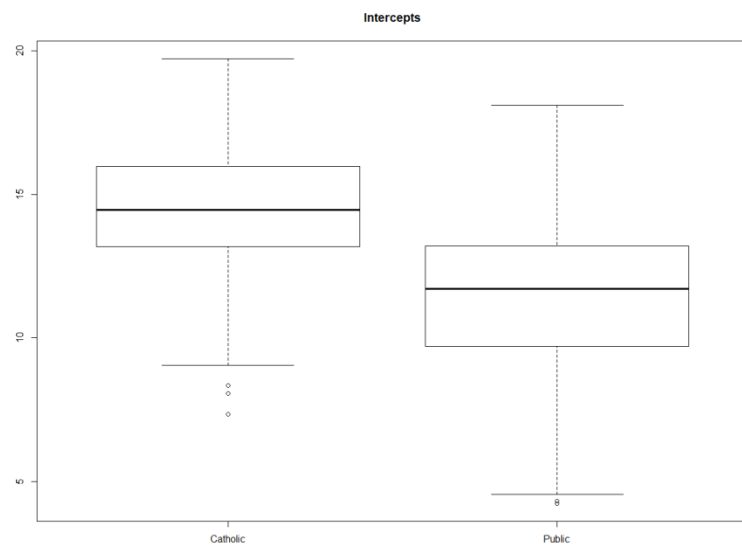
Public



# Интервали поверења



# Разлика у коефицијентима



# Једначина нашег модела

- \*  $MathAch_{ij} = \alpha_{0i} + \alpha_{1i}Cses_{ij} + \varepsilon_{ij}$
- \*  $\alpha_{0i} = \gamma_{00} + \gamma_{01}Meanses_i + \gamma_{02}Sector_i + u_{0i}$
- \*  $\alpha_{1i} = \gamma_{10} + \gamma_{11}Meanses_i + \gamma_{12}Sector_i + u_{1i}$
- \*  $MathAch_{ij} = \gamma_{00} + \gamma_{01}Meanses_i + \gamma_{02}Sector_i + \gamma_{10}Cses_{ij} + \gamma_{11}Meanses_iCses_{ij} + \gamma_{12}Sector_iCses_{ij} + u_{0i} + u_{1i}Cses_{ij} + \varepsilon_{ij}$



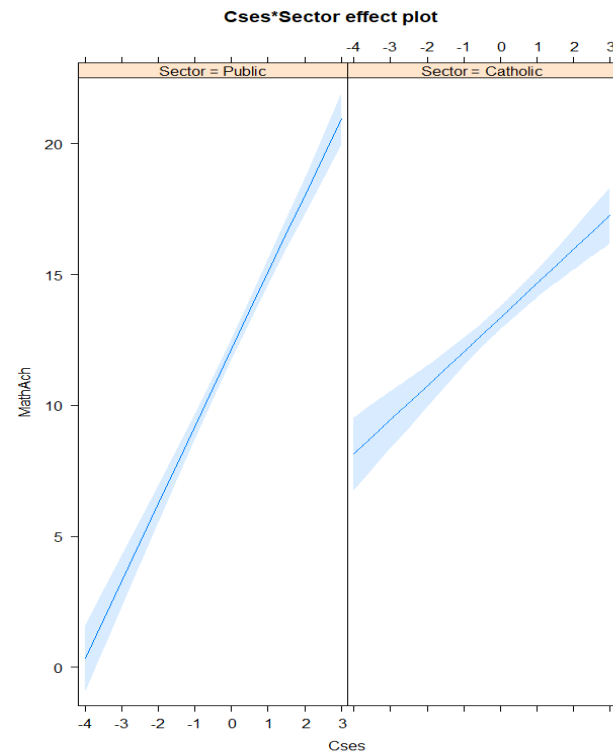
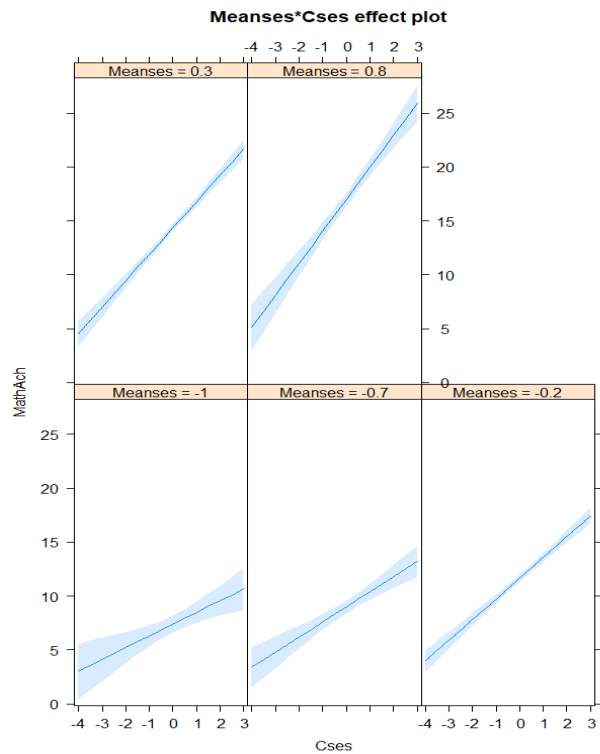
# Функција лме

- \* `Mathlme1 <- lme(MathAch ~ Meanses*Cses +  
Sector*Cses, random = ~ Cses | School, data =Math)  
summary(Mathlme1)`

```
summary(Mathlme1)

## Linear mixed-effects model fit by REML
## Data: Math
##      AIC      BIC    logLik
## 46523.66 46592.45 -23251.83
##
## Random effects:
## Formula: ~Cses | School
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 1.5426082 (Intr)
## Cses        0.3181929 0.391
## Residual    6.0597961
##
## Fixed effects: MathAch ~ Meanses * Cses + Sector * Cses
##                               Value Std.Error   DF t-value p-value
## (Intercept)       12.127931 0.1992913 7022 60.85529 0e+00
## Meanses           5.332875 0.3691672  157 14.44569 0e+00
## Cses              2.945041 0.1556003 7022 18.92696 0e+00
## SectorCatholic    1.226579 0.3062723  157  4.00486 1e-04
## Meanses:Cses      1.039230 0.2988967 7022  3.47689 5e-04
## Cses:SectorCatholic -1.642674 0.2397796 7022 -6.85077 0e+00
## Correlation:
##                               (Intr) Meanss Cses   SctrCt Mnss:C
## Meanses                     0.256
## Cses                        0.075  0.019
## SectorCatholic              -0.699 -0.356 -0.053
## Meanses:Cses                0.019  0.074  0.293 -0.026
## Cses:SectorCatholic         -0.052 -0.027 -0.696  0.077 -0.351
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.15926142 -0.72318922  0.01704599  0.75445035  2.95822019
##
## Number of Observations: 7185
## Number of Groups: 160
```

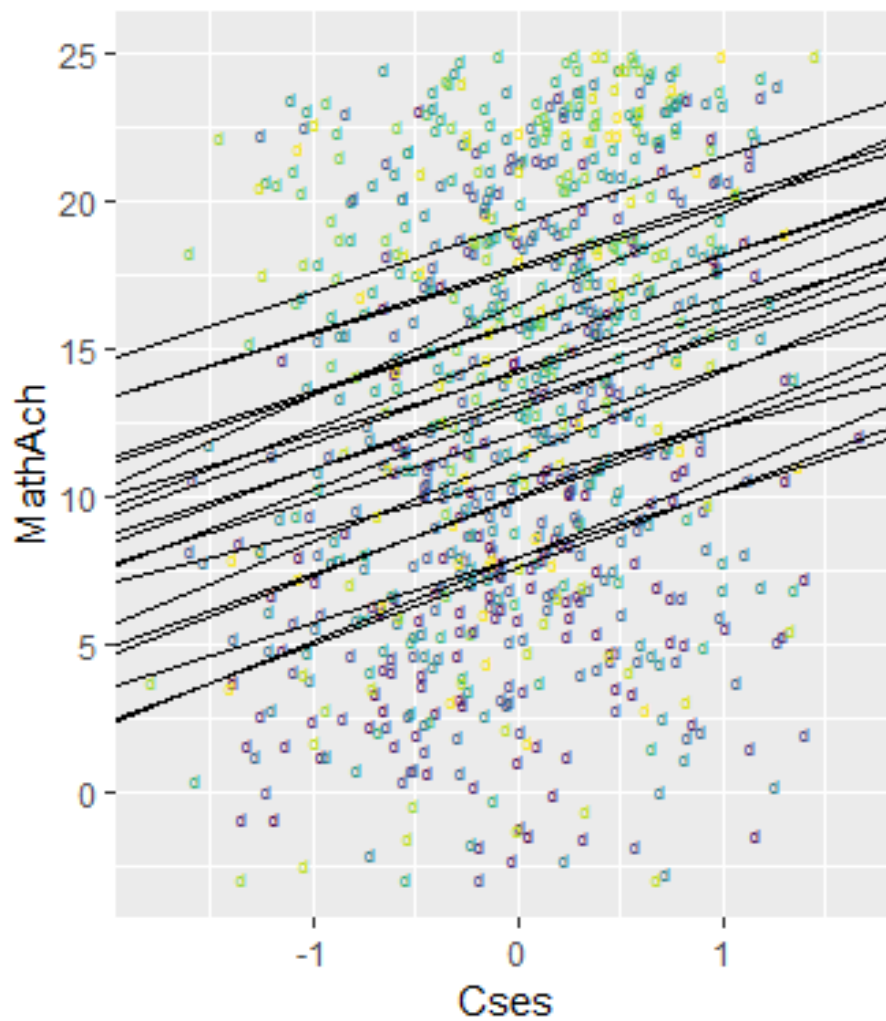
# effects



# Проблеми

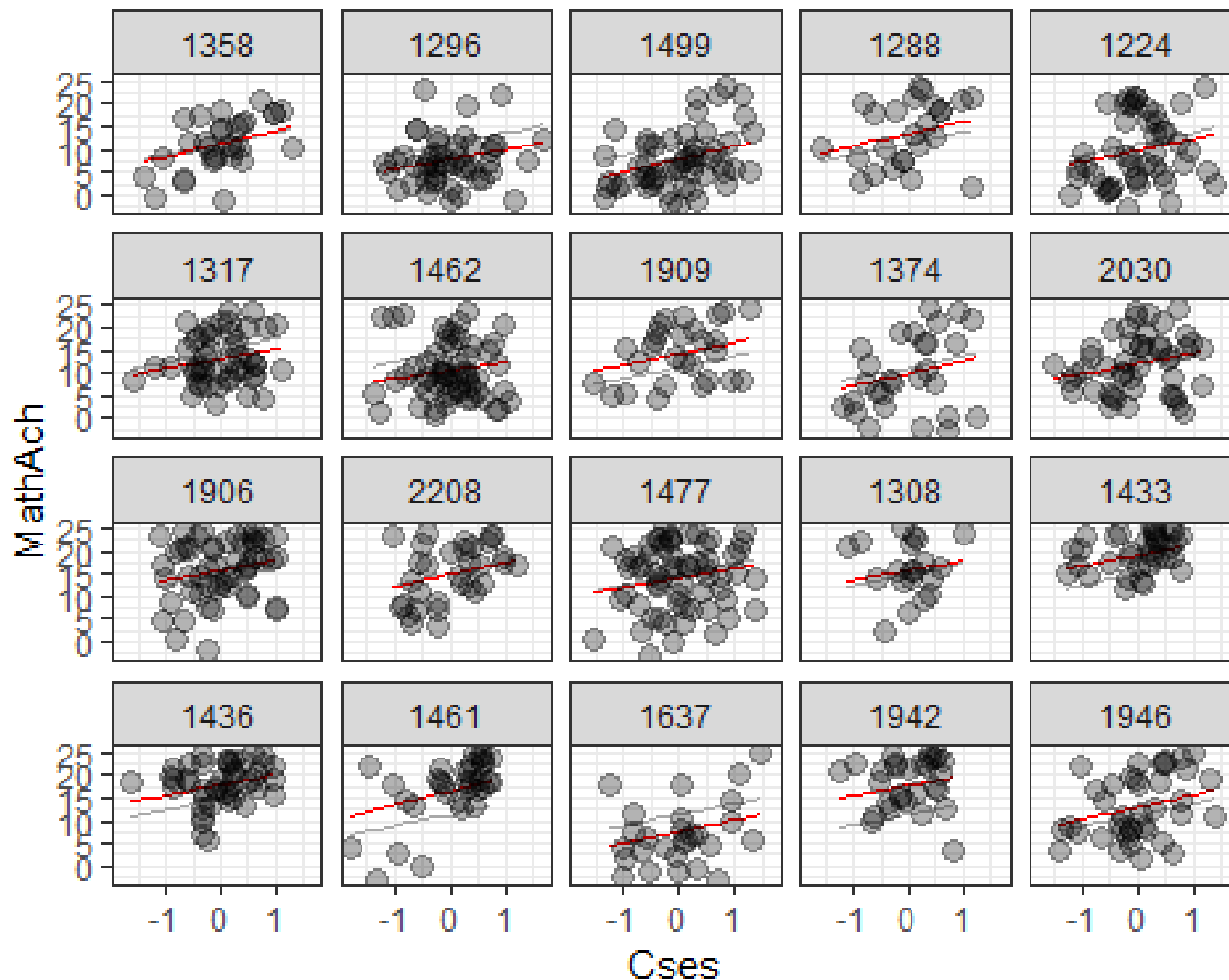
- \* игнорисања важности ефекта групе и претпоставка о независности
- \* са „cross-level“ ефектима

# Различите школе



1499
1288
1224
1317
1462
1909
1374
2030
1906
2208
1477
1308
1433
1436
1461
1637
1942

# Линеарни вс. Хијерархијски



# Предвиђања

- \* ## singular fit
- \* Mlm<- **lm**(MathAch ~ Cses + Sector, train)  
global\_pred<- **predict**(Mlm, newdata=test)  
global\_MSE <- **mean**((test\$MathAch - global\_pred)^2)  
global\_MSE
- \* ## [1] 38.38926
- \* mlm\_pred <-**predict**(Mlmer, newdata=test, allow.new.levels = TRUE)  
mlm\_MSE<- **mean**((test\$MathAch - mlm\_pred)^2)  
mlm\_MSE
- \* ## [1] 33.72613