MULTILEVEL 2 - Data set: IMM10

INTRODUZIONE

In questo data set sono contenuti i dati riferiti a 260 studenti dipartiti in 10 scuole con le seguenti variabili:

1. HOMEWORK: numero di ore settimanali implegate per svolgere i compiti di matematica
2. MATH: punteggio conseguito nel test di matematica
3. MEANSES: media dello stato socio-economico degli studenti nelle singole scuole
4. PARENTED: grado di educazione dei genitori
5. PERCMIN: presenza di minoranze (in percentuale) nelle singole scuole
6. PUBLIC: scuola pubblica (1) o privata (0)
7. RACE: razza dello studente. 1=asiatico, 2=ispanico, 3=di colore, 4=bianco, 5=nativo americano
8. RATIO: rapporto tra numero di alunni e numero di insegnanti all'interno delle singole scuole
9. REGION: codice identificativo della regione in cui è situata la scuola
10. SCHID: codice identificativo della scuola
11. SCHNUM: scuola frequentata dallo studente
12. SCSIZE: dimensioni della scuola (da 1 a 6)
13. SCTYPE: tipologia della scuola. 1=pubblica, 2=cattolica, 3=privata con altra religione, 4=privata non religiosa
14. SES: status socio-economico dello studente
15. SEX: genere dello studente. 1=maschio, 2=femmina
16. STUID: codice identificativo dello studente
17. URBAN: codice identificativo dell'area in cui è sita la scuola
18. WHITE: lo studente è di razza bianca (1) oppure appartiene ad un altra razza (0)

Variabile dipendente: MATH

Analisi proposte:

1. Statistiche descrittive
2. Analisi multilevel

```r
#-- R CODE
library(car)
library(sjstats)
library(plotrix)
library(sjPlot)
library(sjmisc)
library(lme4)
library(pander)
library(car)
library(olsrr)
library(systemfit)
library(het.test)
panderOptions('knitr.auto.asis', FALSE)

#-- White test function
white.test <- function(lmod,data=d){
  u2 <- lmod$residuals^2
  y <- fitted(lmod)
  Ru2 <- summary(lm(u2 ~ y + I(y^2)))$r.squared
  LM <- nrow(data)*Ru2
```

```r
  p.value <- 1-pchisq(LM, 2)
  data.frame("Test statistic"=LM,"P value"=p.value)
}

#-- funzione per ottenere osservazioni outlier univariate
FIND_EXTREME_OBSERVARION <- function(x,sd_factor=2){
  which(x>mean(x)+sd_factor*sd(x) | x<mean(x)-sd_factor*sd(x))
}

#-- import dei dati
ABSOLUTE_PATH <- "C:\\Users\\sbarberis\\Dropbox\\MODELLI STATISTICI"
d <- read.csv(paste0(ABSOLUTE_PATH,"\\esercizi (3) copia\\2.multilevel\\imm10.csv"),sep=";")

#-- Fisso la decima scuola come riferimento
d$schnum <- factor(d$schnum)
contrasts(d$schnum) <- contr.treatment(levels(d$schnum),base=which(levels(d$schnum) == '10'))


#-- vettore di variabili numeriche presenti nei dati
VAR_NUMERIC <- c("homework","percmin","math","meanses")

#-- print delle prime 6 righe del dataset
pander(head(d),big.mark=",")
```

Table 1: Table continues below

schid	stuid	ses	meanses	homework	white	parented	public	ratio
7,472	3	-0.13	-0.4826	1	1	2	1	19
7,472	8	-0.39	-0.4826	0	1	2	1	19
7,472	13	-0.8	-0.4826	0	1	2	1	19
7,472	17	-0.72	-0.4826	1	1	2	1	19
7,472	27	-0.74	-0.4826	2	1	2	1	19
7,472	28	-0.58	-0.4826	1	1	2	1	19

percmin	math	sex	race	sctype	scsize	urban	region	schnum
0	48	2	4	1	3	2	2	1
0	48	1	4	1	3	2	2	1
0	53	1	4	1	3	2	2	1
0	42	1	4	1	3	2	2	1
0	43	2	4	1	3	2	2	1
0	57	2	4	1	3	2	2	1

STATISTICHE DESCRITTIVE

Si propongono la matrice di correlazione tra le variabili e alcune descrittive di base.

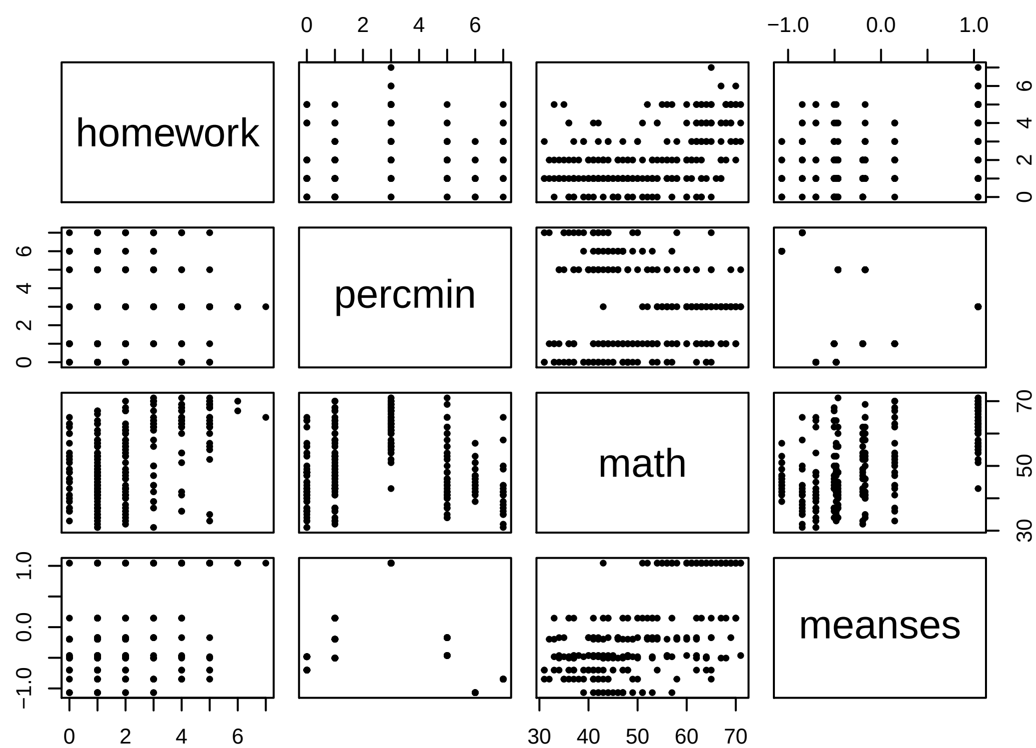
```r
#-- R CODE
pander(summary(d[,VAR_NUMERIC])) #-- statistiche descrittive
```

homework	percmin	math	meanses
Min. :0.000	Min. :0.000	Min. :31.0	Min. :-1.06850
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:42.0	1st Qu.:-0.50450
Median :1.000	Median :3.000	Median :49.5	Median :-0.19682
Mean :2.023	Mean :2.835	Mean :51.3	Mean :-0.07331
3rd Qu.:3.000	3rd Qu.:5.000	3rd Qu.:62.0	3rd Qu.: 1.04463
Max. :7.000	Max. :7.000	Max. :71.0	Max. : 1.04463

```r
pander(cor(d[,VAR_NUMERIC])) #-- matrice di correlazione
```

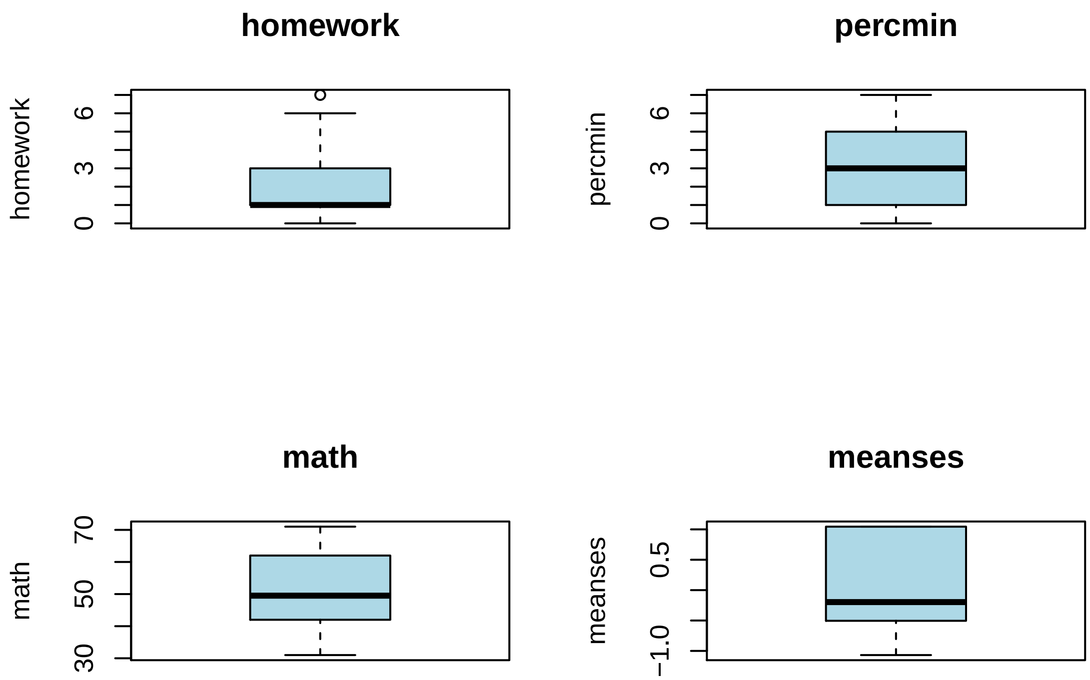
	homework	percmin	math	meanses
homework	1	0.1034	0.497	0.4474
percmin	0.1034	1	-0.04592	-0.1416
math	0.497	-0.04592	1	0.6442
meanses	0.4474	-0.1416	0.6442	1

```r
plot(d[,VAR_NUMERIC],pch=19,cex=.5) #-- scatter plot multivariato
```

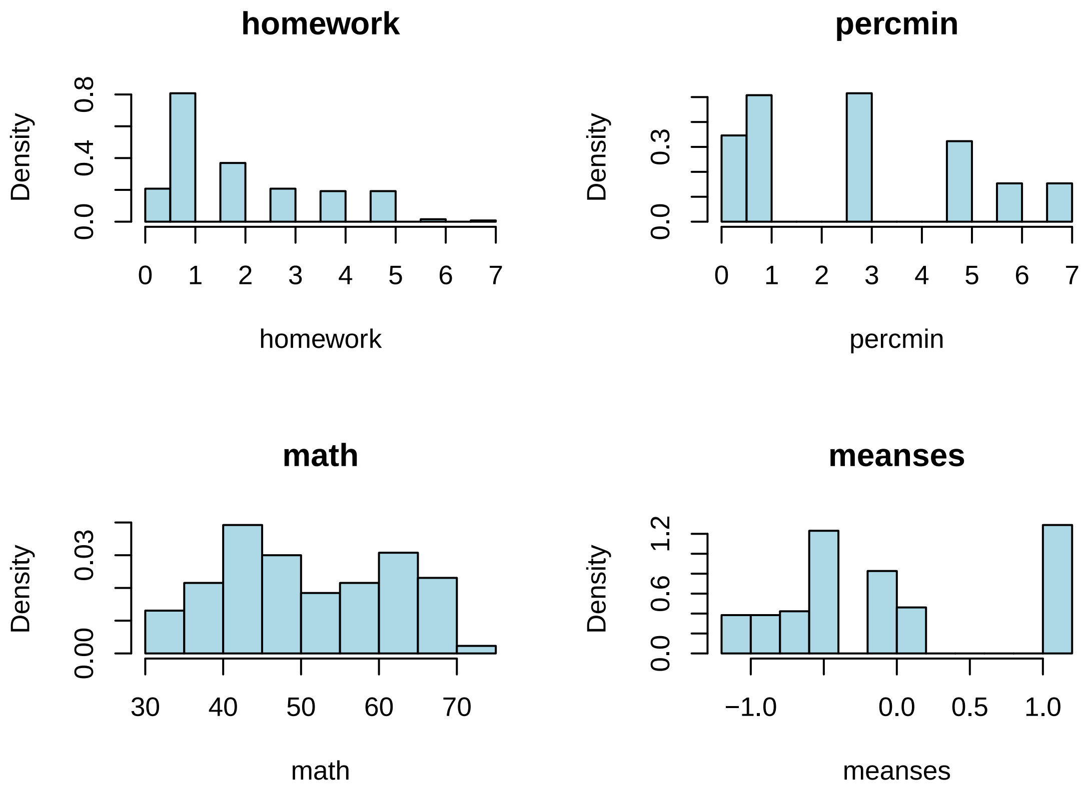



```r
par(mfrow=c(2,2))
for(i in VAR_NUMERIC){
```

```
  boxplot(d[,i],main=i,col="lightblue",ylab=i)
}
```

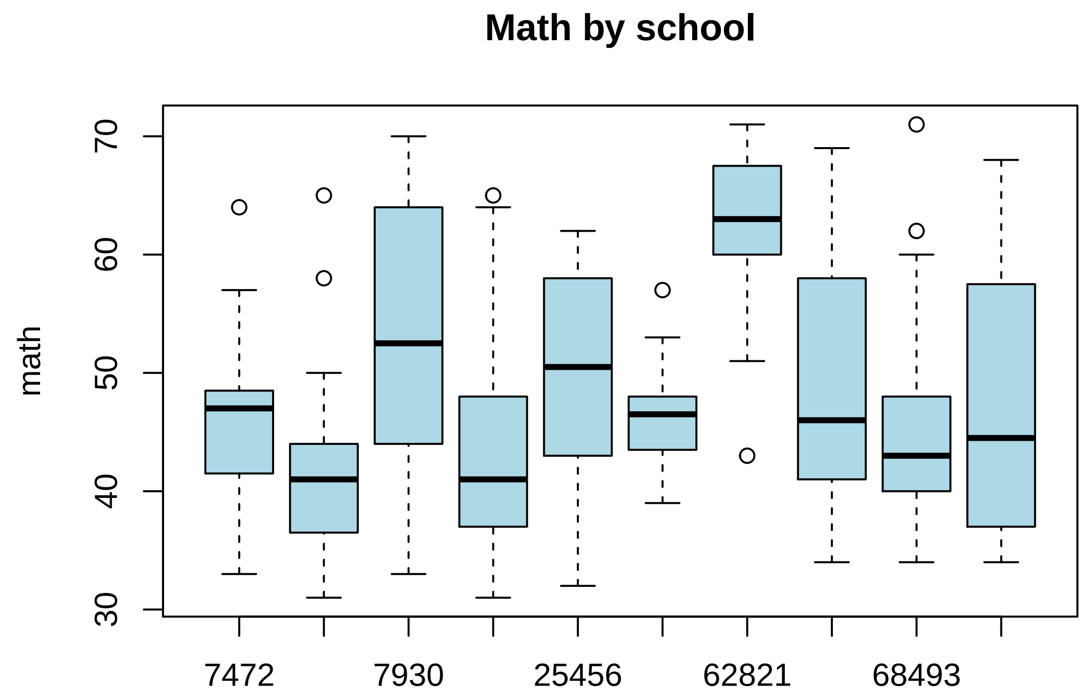



```
par(mfrow=c(2,2))
for(i in VAR_NUMERIC){
  hist(d[,i],main=i,col="lightblue",xlab=i,freq=F)
}
```



Proponiamo ora il box plot per la variabile "math" che sarà la variabile dipendente nel modello Multilevel:

```
#-- R CODE
boxplot(d$math~d$schid,main="Math by school",col="lightblue",ylab="math")
```



ANALISI DELLA VARIANZA (EFFETTI FISSI)

Si consideri ora innanzitutto una analisi della varianza a effetti fissi.

```
#-- R CODE
mod1 <- lm(math ~ schnum,d)
summary(mod1)
```

```
##
## Call:
## lm(formula = math ~ schnum, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.2500  -5.7052  -0.6423   4.4909  24.6667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   47.850      1.902  25.159  < 2e-16 ***
## schnum1       -2.111      2.601  -0.812   0.4177
## schnum2       -5.700      2.690  -2.119   0.0351 *
## schnum3        5.400      2.575   2.097   0.0370 *
## schnum4       -4.305      2.628  -1.638   0.1027
## schnum5        2.014      2.628   0.766   0.4442
```

```
## schnum6       -1.450      2.690  -0.539   0.5903
## schnum7       14.971      2.167   6.908 4.06e-11 ***
## schnum8        1.817      2.657   0.684   0.4949
## schnum9       -1.517      2.657  -0.571   0.5687
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.506 on 250 degrees of freedom
## Multiple R-squared:  0.4369, Adjusted R-squared:  0.4166
## F-statistic: 21.55 on 9 and 250 DF,  p-value: < 2.2e-16
```

```
pander(anova(mod1),big.mark=",")
```

```
##
## ---------------------------------------------------------------
##             Df    Sum Sq   Mean Sq   F value    Pr(>F)
## ---------------- ----- -------- --------- --------- -----------
##   **schnum**      9    14,031    1,559     21.55    7.661e-27
##
##  **Residuals**   250   18,086    72.34      NA         NA
## ---------------------------------------------------------------
##
## Table: Analysis of Variance Table
```

Viene respinta l'ipotesi che le medie dei gruppi siano tutte uguali. Analizzandole e avendo come riferimento la media del gruppo 10 si osserva ad esempio che la scuola 7 ha un punteggio in math di 14.97 punti più elevato rispetto alla scuola 10 stessa, mentre la scuola 2 avrà un punteggio inferiore in media di 5.7. Si passa ora al modello empty.

REGRESSIONE MULTILEVEL: Empty Model

```
#-- R CODE
mod1 <- lmer(math ~ (1| schnum),d,REML=T)
summary(mod1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ (1 | schnum)
##    Data: d
##
## REML criterion at convergence: 1871.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.34027 -0.65554 -0.08304  0.54211  2.87452
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schnum   (Intercept) 34.01    5.832
##  Residual             72.26    8.500
## Number of obs: 260, groups:  schnum, 10
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   48.861      1.927   25.35
```

```r
pander(Anova(mod1, type="III"),big.mark=",")
```

```
##
## ---------------------------------------------
##               Chisq   Df   Pr(>Chisq)
## ----------------- ------- ---- ------------
##  **(Intercept)**   642.7   1    8.76e-142
## ---------------------------------------------
##
## Table: Analysis of Deviance Table (Type III Wald chisquare tests)
```

```r
mod1_null <- lm(math ~ 1,d)
pander(anova(mod1,mod1_null),big.mark=",")
```

```
##
## ---------------------------------------------------------------------------------------
##             Df    AIC     BIC     logLik    deviance   Chisq   Chi Df   Pr(>Chisq)
## ---------------- ---- ------- ------- --------- ---------- ------- -------- ------------
##  **mod1_null**   2    1,994   2,001   -995.1     1,990      NA       NA         NA
##
##    **mod1**      3    1,881   1,891   -937.4     1,875     115.3     1      6.613e-27
## ---------------------------------------------------------------------------------------
##
## Table: Data: d
```

```r
pander(data.frame("ICC"=icc(mod1)),big.mark=",") #-- ICC
```

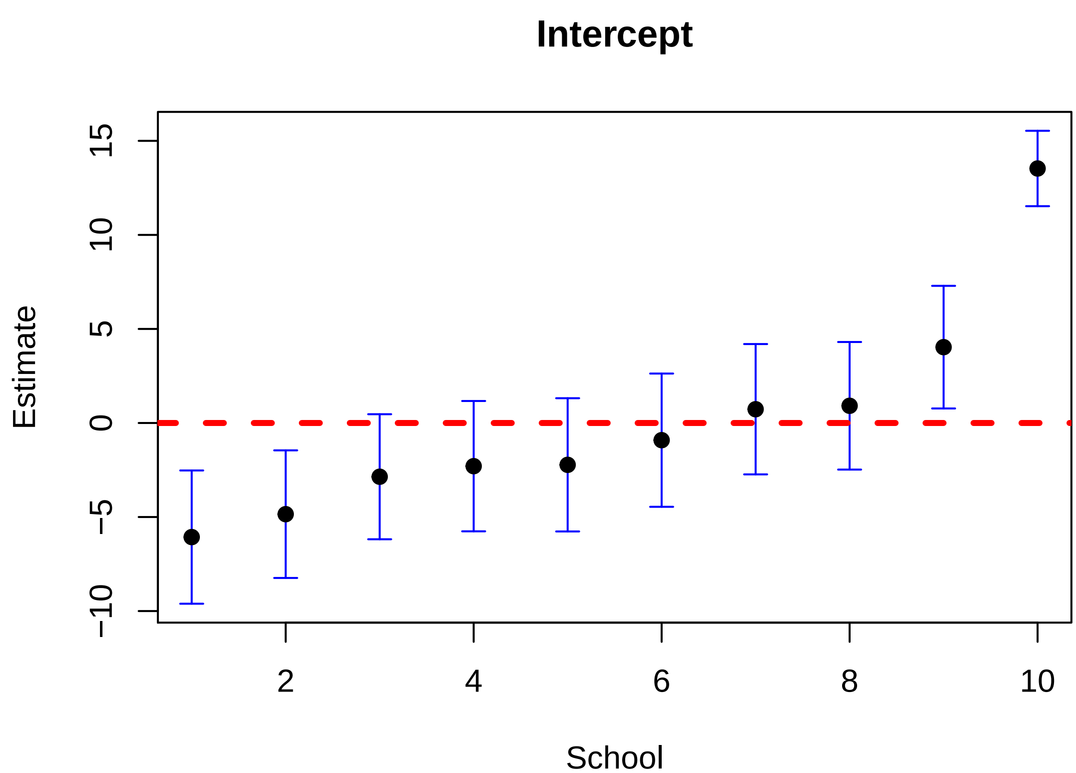
```
##
## ---------------------
##           ICC
## ------------ --------
##  **schnum**   0.3201
## ---------------------
```

```r
res <- sjp.lmer(mod1, type = "re.qq", sort.est = "sort.all",show.values=T,title="T",prnt.plot=F)
res$data$lower <- res$data$y-res$data$ci
res$data$upper <- res$data$y+res$data$ci

pander(res$data[1:10,c("ID","y","upper","lower")])
```

```
##
## ---------------------------------------------------
##                 ID      y       upper    lower
## ------------------- ---- --------- -------- --------
##  **(Intercept)2**    2    -6.067    -2.525   -9.609
##
##  **(Intercept)4**    4    -4.848    -1.456   -8.24
##
##  **(Intercept)1**    1    -2.858    0.4657   -6.182
##
##  **(Intercept)9**    9    -2.296    1.169    -5.76
##
##  **(Intercept)6**    6    -2.225    1.317    -5.767
##
##  **(Intercept)10**   10   -0.9142   2.628    -4.456
##
```

```
##   **(Intercept)8**    8    0.7314    4.196    -2.733
##
##   **(Intercept)5**    5     0.914    4.306    -2.478
##
##   **(Intercept)3**    3     4.032    7.291    0.7721
##
##   **(Intercept)7**    7     13.53    15.53    11.53
## ---------------------------------------------------
```

```
plotCI(1:nrow(res$data),res$data$y,ui=res$data$upper, li=res$data$lower,pch=19,scol="blue",xlab="School
abline(h=mean(res$data$y),col=2,lwd=3,lty=2)
```



E' respinta l'ipotesi che il modello non interpreti i dati e dal rapporto tra varianza spiegata e totale si ricava un coefficiente intraclasse pari a 0.32 che è elevato, e segnala una buona variabilità fra le scuole nei punteggi di matematica. Si propongono poi i valori attesi e gli intervalli di confidenza dei parametri casuali inerenti le singole scuole. Per i parametri casuali il modello postula graduatorie basate su valori attesi e intervalli di confidenza.

Come è noto una scuola A può ritenersi superiore a una scuola B in termini di efficacia solo se l'estremo inferiore dell'intervallo di confidenza di A sia superiore all'estremo superiore dell'intervallo di confidenza di B. Si può notare che 6 scuole su 10 hanno un andamento peggiore rispetto a quello della media generale. Si può verificare che solo l'effetto casuale del gruppo 10 è significativo. Si passa poi a proporre il modello random intercept introducendo prima la variabile esplicativa "homework".

REGRESSIONE MULTILEVEL: Random Intercept

```r
#-- R CODE
mod1 <- lmer(math ~ homework + (1| schnum),d,REML=T)
summary(mod1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ homework + (1 | schnum)
##    Data: d
##
## REML criterion at convergence: 1839.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6060 -0.6872 -0.0244  0.5983  3.3770
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schnum   (Intercept) 25.22    5.022
##  Residual             64.52    8.033
## Number of obs: 260, groups:  schnum, 10
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   44.982      1.803  24.949
## homework       2.207      0.379   5.823
##
## Correlation of Fixed Effects:
##          (Intr)
## homework -0.371
```

```r
pander(Anova(mod1, type="III"),big.mark=",")
```

```
##
## -------------------------------------------
##               Chisq   Df   Pr(>Chisq)
## ----------------- ------- ---- ------------
##  **(Intercept)**   622.4   1    2.201e-137
##
##   **homework**     33.9    1    5.792e-09
## -------------------------------------------
##
## Table: Analysis of Deviance Table (Type III Wald chisquare tests)
```

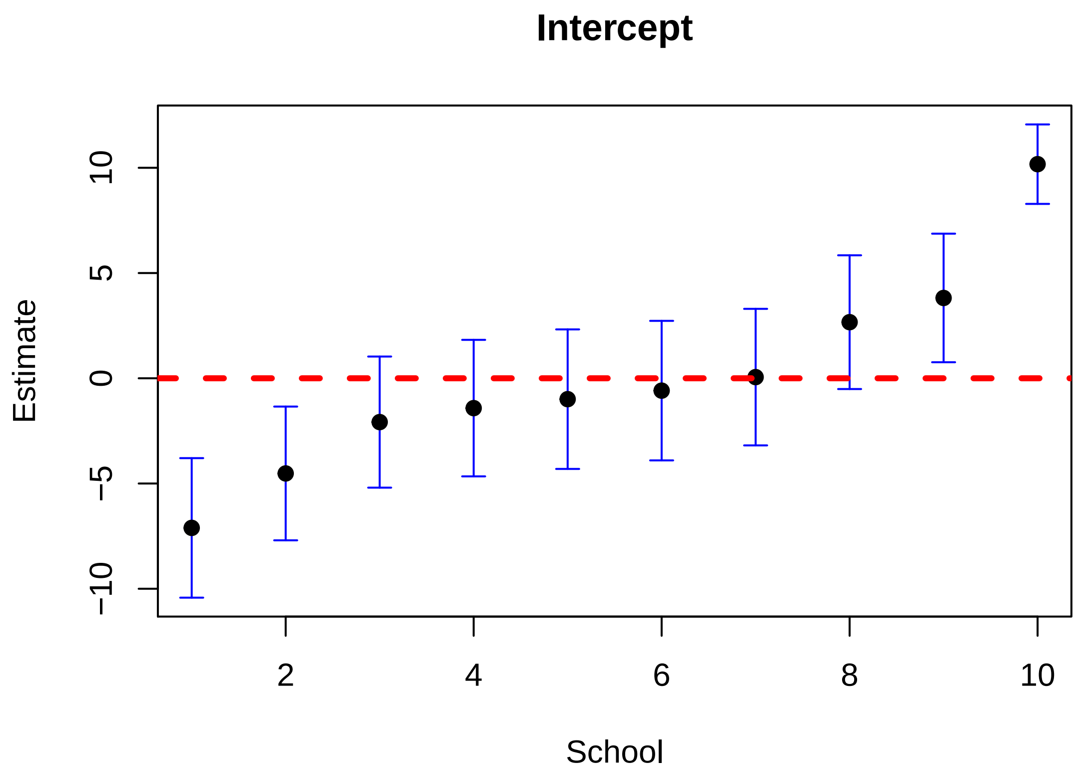
```r
pander(data.frame("ICC"=icc(mod1)),big.mark=",") #-- ICC
```

```
##
## ---------------------
##           ICC
## ------------ --------
##  **schnum**   0.2811
## ---------------------
```

```r
res <- sjp.lmer(mod1, type = "re.qq", sort.est = "sort.all",show.values=T,title="T",prnt.plot=F)
res$data$lower <- res$data$y-res$data$ci
res$data$upper <- res$data$y+res$data$ci

pander(res$data[1:10,c("ID","y","upper","lower")],big.mark=",")
```

```
##
## ---------------------------------------------------
##                 ID      y       upper    lower
## ------------------- ---- --------- -------- --------
##  **(Intercept)2**    2    -7.108    -3.793   -10.42
##
##  **(Intercept)4**    4    -4.522    -1.345   -7.699
##
##  **(Intercept)1**    1    -2.081    1.033    -5.196
##
##  **(Intercept)9**    9    -1.418    1.826    -4.662
##
##  **(Intercept)6**    6    -0.9926   2.322    -4.307
##
##  **(Intercept)10**   10   -0.5873   2.728    -3.902
##
##  **(Intercept)8**    8    0.05477   3.299    -3.189
##
##  **(Intercept)5**    5     2.666    5.843    -0.511
##
##  **(Intercept)3**    3     3.816    6.871    0.7609
##
##  **(Intercept)7**    7     10.17    12.06    8.285
## ---------------------------------------------------
```

```
plotCI(1:nrow(res$data),res$data$y,ui=res$data$upper, li=res$data$lower,pch=19,scol="blue",xlab="School
abline(h=mean(res$data$y),col=2,lwd=3,lty=2)
```



Il coefficiente di correlazione intraclasse si abbassa di pochissimo (0.28) in quanto si abbassano in uguale proporzione varianza spiegata e residua. Il modello interpreta bene i dati e la variabile "homework" risulta altresì significativa. Anche il test di 3° tipo degli effetti fissi conferma questa significatività.

Si propongono i valori attesi e gli intervalli di confidenza dei parametri casuali inerenti i gruppi. Si vede come il ranking muti in modo rilevante al caso empty. Ciò mostra che la diversa distribuzione fra le scuole della variabile "homework" è all'origine di parte della variabilità di "math" attribuito in prima istanza nel modello empty alla efficacia delle scuole.

Tenere conto di questo non solo modifica l'efficacia complessiva delle scuole ma anche l'efficacia relativa di ogni scuola rispetto ad altre. Si può verificare che solo l'effetto casuale del gruppo 7 è significativo e anche in questo caso quindi la situazione cambia radicalmente rispetto al modello empty.

Si aggiunge ora nel modello mixed anche la variabile esplicativa "ses".

```
#-- R CODE
mod1 <- lmer(math ~ homework + ses + (1| schnum),d,REML=T)
summary(mod1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ homework + ses + (1 | schnum)
##    Data: d
##
## REML criterion at convergence: 1815
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.5104 -0.6830  0.0149  0.6194  3.4263
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schnum   (Intercept)  9.609   3.100
##  Residual             60.658   7.788
## Number of obs: 260, groups:  schnum, 10
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  46.6453     1.3193   35.36
## homework      1.9734     0.3673    5.37
## ses           3.8004     0.7082    5.37
##
## Correlation of Fixed Effects:
##          (Intr) homwrk
## homework -0.523
## ses       0.241 -0.166
```

```
pander(Anova(mod1, type="III"),big.mark=",")
```

```
##
## -------------------------------------------
##               Chisq   Df   Pr(>Chisq)
## ----------------- ------- ---- ------------
##  **(Intercept)**   1,250   1    7.768e-274
##
##   **homework**     28.86   1    7.766e-08
##
##      **ses**       28.8    1    8.032e-08
## -------------------------------------------
##
## Table: Analysis of Deviance Table (Type III Wald chisquare tests)
```

```
pander(data.frame("ICC"=icc(mod1)),big.mark=",") #-- ICC
```

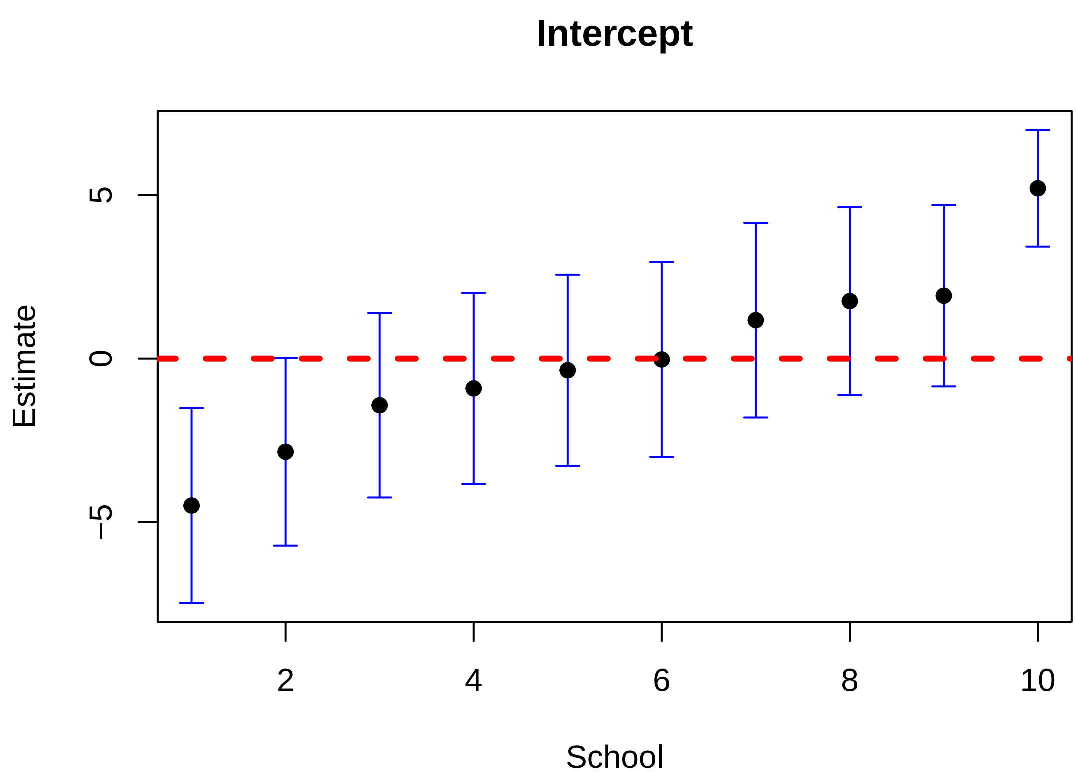
```
##
## ---------------------
##           ICC
## ------------ --------
##  **schnum**   0.1368
## ---------------------
```

```
res <- sjp.lmer(mod1, type = "re.qq", sort.est = "sort.all",show.values=T,title="T",prnt.plot=F)
res$data$lower <- res$data$y-res$data$ci
res$data$upper <- res$data$y+res$data$ci

pander(res$data[1:10,c("ID","y","upper","lower")],big.mark=",")
```

```
##
## -------------------------------------------------------
##                 ID       y        upper     lower
## ------------------- ---- ---------- --------- ---------
##  **(Intercept)2**    2     -4.492    -1.516    -7.468
##
##  **(Intercept)4**    4     -2.848    0.02074   -5.717
```

```
##
##  **(Intercept)1**    1     -1.426     1.393    -4.246
##
##  **(Intercept)9**    9    -0.9105     2.01     -3.831
##
##  **(Intercept)8**    8    -0.3566     2.564    -3.277
##
##  **(Intercept)10**   10   -0.02698    2.949    -3.003
##
##  **(Intercept)6**    6     1.175      4.151    -1.801
##
##  **(Intercept)5**    5     1.758      4.627    -1.111
##
##  **(Intercept)3**    3     1.922      4.695    -0.8504
##
##  **(Intercept)7**    7     5.206      6.989     3.423
## -------------------------------------------------------
```

```r
plotCI(1:nrow(res$data),res$data$y,ui=res$data$upper, li=res$data$lower,pch=19,scol="blue",xlab="School
abline(h=mean(res$data$y),col=2,lwd=3,lty=2)
```



Il coefficiente di correlazione intraclasse si dimezza (0.136) perché diminuisce la varianza spiegata molto più che la varianza complessiva a segnalare che la variabile “ses”, molto più che “homework” cattura la variabilità di “math”. In altre parole molta della variabilità che sembrava doversi attribuire alle scuole è invece dovuta alla diversa distribuzione fra le scuole della variabile “ses”.

La diversa composizione socio-economica delle scuole cattura quindi una parte della variabilità della variabile dipendente e non può essere quindi attribuita alla diversa efficacia delle scuole. Per il resto anche questo modello è significativo come anche le variabili esplicative. Il test di 3° tipo degli effetti fissi conferma la significatività di queste variabili.

Si osserva come sia i valori attesi che gli intervalli di confidenza dei gruppi mutano per la diversa influenza dello stato-socioeconomico nelle diverse scuole. Tenere conto di questo non solo modifica l'efficacia complessiva delle scuole ma anche l'efficacia relativa di ogni scuola rispetto ad altre. Si può verificare che solo l'effetto casuale del gruppo 7 rimane significativo anche se in misura minore che nel caso con variabile esplicativa solo "homework" e in parte diviene significativo l'effetto 2 e 4. Si passa ora a un modello random effect in cui anche il parametro relativo a "homework" è di tipo casuale.

REGRESSIONE MULTILEVEL: Random Slope

```
#-- R CODE
mod1 <- lmer(math ~ homework + ses + (homework| schnum),d,REML=T)
summary(mod1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ homework + ses + (homework | schnum)
##    Data: d
##
## REML criterion at convergence: 1745.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.46818 -0.67278 -0.00633  0.64581  2.63591
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  schnum   (Intercept) 55.17    7.428
##           homework    20.10    4.483    -0.88
##  Residual             41.27    6.425
## Number of obs: 260, groups:  schnum, 10
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  46.1123     2.4840  18.564
## homework      1.8198     1.4742   1.234
## ses           2.7601     0.6099   4.525
##
## Correlation of Fixed Effects:
##          (Intr) homwrk
## homework -0.873
## ses       0.118 -0.036
```

```
pander(Anova(mod1, type="III"),big.mark=",")
```

```
##
## -------------------------------------------
##               Chisq   Df   Pr(>Chisq)
## ----------------- ------- ---- ------------
##  **(Intercept)**   344.6   1    6.331e-77
##
##   **homework**     1.524   1      0.217
```

```
##
##      **ses**       20.48   1    6.035e-06
## --------------------------------------------
##
## Table: Analysis of Deviance Table (Type III Wald chisquare tests)
```

```r
res <- sjp.lmer(mod1, type = "re.qq", sort.est = "sort.all",show.values=T,title="T",prnt.plot=F)
res$data$lower <- res$data$y-res$data$ci
res$data$upper <- res$data$y+res$data$ci

res_int <- subset(res$data,ind=="(Intercept)")
res_hw <- subset(res$data,ind=="homework")

pander(res_int[1:10,c("ID","y","upper","lower")])
```

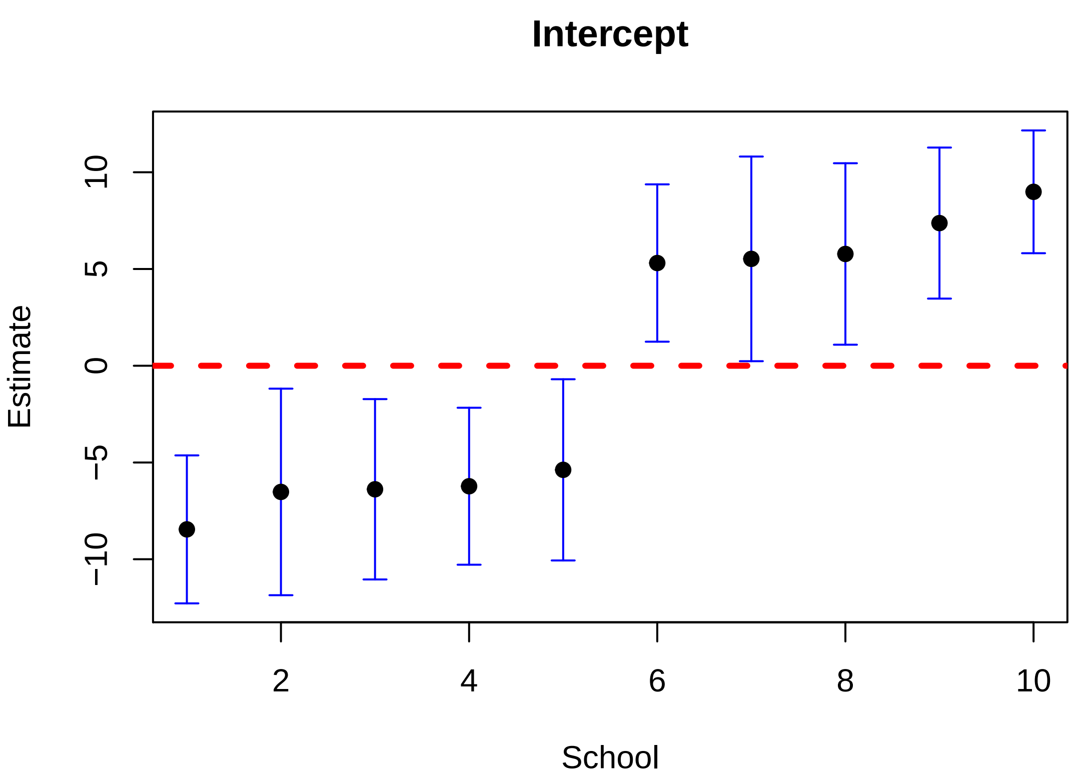
```
##
## ----------------------------------------------------
##                  ID     y       upper    lower
## ------------------- ---- -------- --------- --------
##  **(Intercept)4**    4    -8.457   -4.633    -12.28
##
##  **(Intercept)8**    8    -6.521   -1.184    -11.86
##
##  **(Intercept)3**    3    -6.385   -1.725    -11.04
##
##  **(Intercept)10**   10   -6.228   -2.172    -10.28
##
##  **(Intercept)9**    9    -5.38    -0.6978   -10.06
##
##  **(Intercept)1**    1    5.308     9.374    1.243
##
##  **(Intercept)2**    2    5.524     10.81    0.2355
##
##  **(Intercept)6**    6    5.777     10.47    1.086
##
##  **(Intercept)5**    5    7.373     11.28     3.47
##
##  **(Intercept)7**    7    8.988     12.16    5.816
## ----------------------------------------------------
```

```r
pander(res_hw[1:10,c("ID","y","upper","lower")])
```

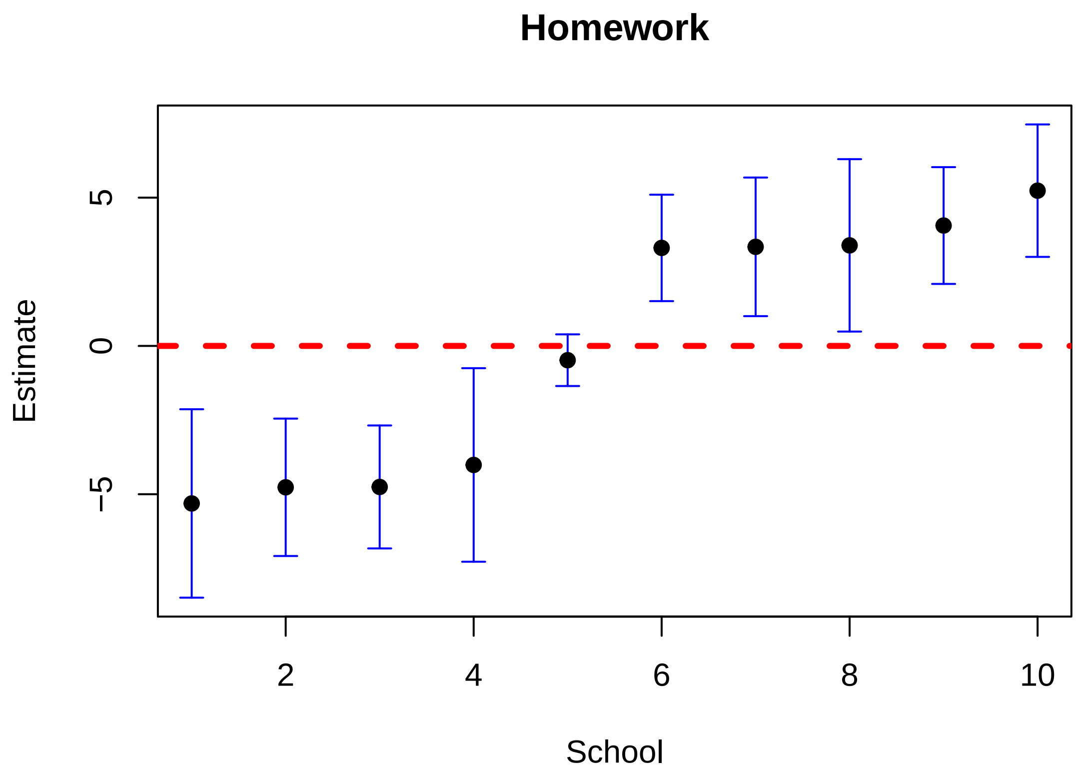
```
##
## --------------------------------------------------
##               ID      y       upper    lower
## ---------------- ---- --------- --------- --------
##  **homework5**    5    -5.312    -2.135    -8.489
##
##  **homework1**    1    -4.767    -2.451    -7.084
##
##  **homework2**    2    -4.756    -2.683    -6.829
##
##  **homework6**    6    -4.013    -0.7504   -7.276
##
##  **homework7**    7    -0.4815   0.3895    -1.352
```

```
##
##   **homework4**    4     3.305      5.1       1.51
##
##   **homework8**    8     3.341     5.678     1.004
##
##   **homework9**    9     3.39      6.297     0.4837
##
##   **homework10**   10    4.059     6.028      2.09
##
##   **homework3**    3     5.235     7.468     3.002
## --------------------------------------------------
```

```
plotCI(1:nrow(res_int),res_int$y,ui=res_int$upper, li=res_int$lower,pch=19,scol="blue",xlab="School",yla
abline(h=mean(res_int$y),col=2,lwd=3,lty=2)
```




```
plotCI(1:nrow(res_hw),res_hw$y,ui=res_hw$upper, li=res_hw$lower,pch=19,scol="blue",xlab="School",ylab="
abline(h=mean(res_hw$y),col=2,lwd=3,lty=2)
```



Il coefficiente intraclasse non può più essere calcolato nel modo semplice precedente perché si deve tener conto della correlazione tra effetti casuali relativi a "homework" e alle scuole nel loro complesso. Gli effetti casuali complessivi relativi all'efficacia delle scuole nel loro complesso e "homework" sono significativi anche se con un p-value non molto basso. La correlazione tra effetti relativi alle scuole e a "homework" è negativa. Inoltre il modello interpreta bene i dati ma la parte fissa della variabile "homework" non è più significativa. Questo risultato è confermato anche dal test 3 degli effetti fissi. Per ciò che concerne gli effetti casuali di "homework" diversi tra scuola e scuola non risulta significativo per nessun p-value eccetto quello relativo alla scuola 7; per le intercette inerenti l'efficacia relativa di ogni scuola solo quello relativo alla scuola 2.

Si aggiunge ora la variabile SES di primo livello e ratio di secondo.

```
#-- R CODE
mod1 <- lmer(math ~ homework + ses + ratio + (homework| schnum),d,REML=T)
summary(mod1)
```

Linear mixed model fit by REML ['lmerMod'] Formula: math ~ homework + ses + ratio + (homework | schnum) Data: d

REML criterion at convergence: 1742.2

Scaled residuals: Min 1Q Median 3Q Max -2.51862 -0.67131 0.01126 0.64162 2.69150

Random effects: Groups Name Variance Std.Dev. Corr schnum (Intercept) 43.22 6.574
homework 19.27 4.389 -0.90 Residual 41.34 6.429
Number of obs: 260, groups: schnum, 10

Fixed effects: Estimate Std. Error t value (Intercept) 55.2100 4.7507 11.621 homework 1.8410 1.4457 1.273 ses 2.7227 0.6077 4.481 ratio -0.5889 0.2753 -2.139

Correlation of Fixed Effects: (Intr) homwrk ses
homework -0.400
ses -0.077 -0.038
ratio -0.883 -0.017 0.154

```
pander(Anova(mod1, type="III"),big.mark=",")
```

Table 5: Analysis of Deviance Table (Type III Wald chisquare tests)

	Chisq	Df	Pr(>Chisq)
(Intercept)	135.1	1	3.209e-31
homework	1.622	1	0.2029
ses	20.08	1	7.444e-06
ratio	4.575	1	0.03244

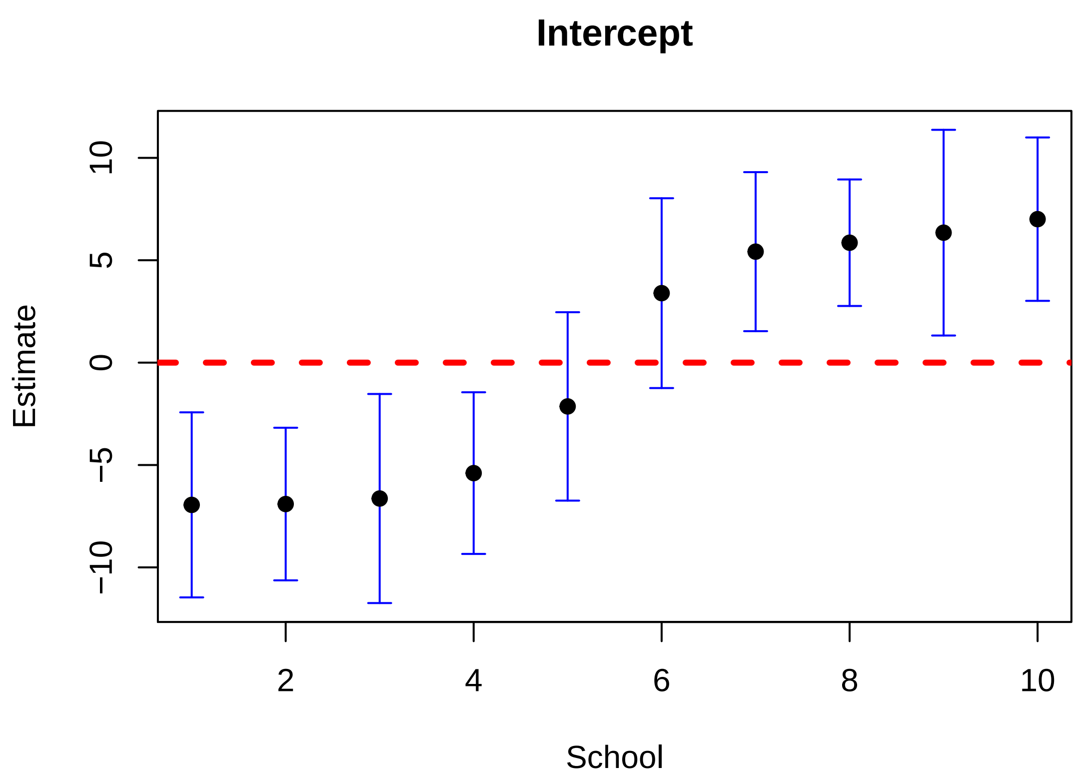
```
res <- sjp.lmer(mod1, type = "re.qq", sort.est = "sort.all",show.values=T,title="T",prnt.plot=F)
res$data$lower <- res$data$y-res$data$ci
res$data$upper <- res$data$y+res$data$ci

res_int <- subset(res$data,ind=="(Intercept)")
res_hw <- subset(res$data,ind=="homework")

pander(res_int[1:10,c("ID","y","upper","lower")],big.mark=",")
```

	ID	y	upper	lower
(Intercept)3	3	-6.947	-2.428	-11.47
(Intercept)4	4	-6.906	-3.18	-10.63
(Intercept)8	8	-6.637	-1.532	-11.74
(Intercept)10	10	-5.395	-1.448	-9.343
(Intercept)9	9	-2.138	2.462	-6.739
(Intercept)6	6	3.393	8.028	-1.241
(Intercept)5	5	5.42	9.303	1.537
(Intercept)7	7	5.856	8.946	2.767
(Intercept)2	2	6.347	11.37	1.324
(Intercept)1	1	7.007	11	3.018

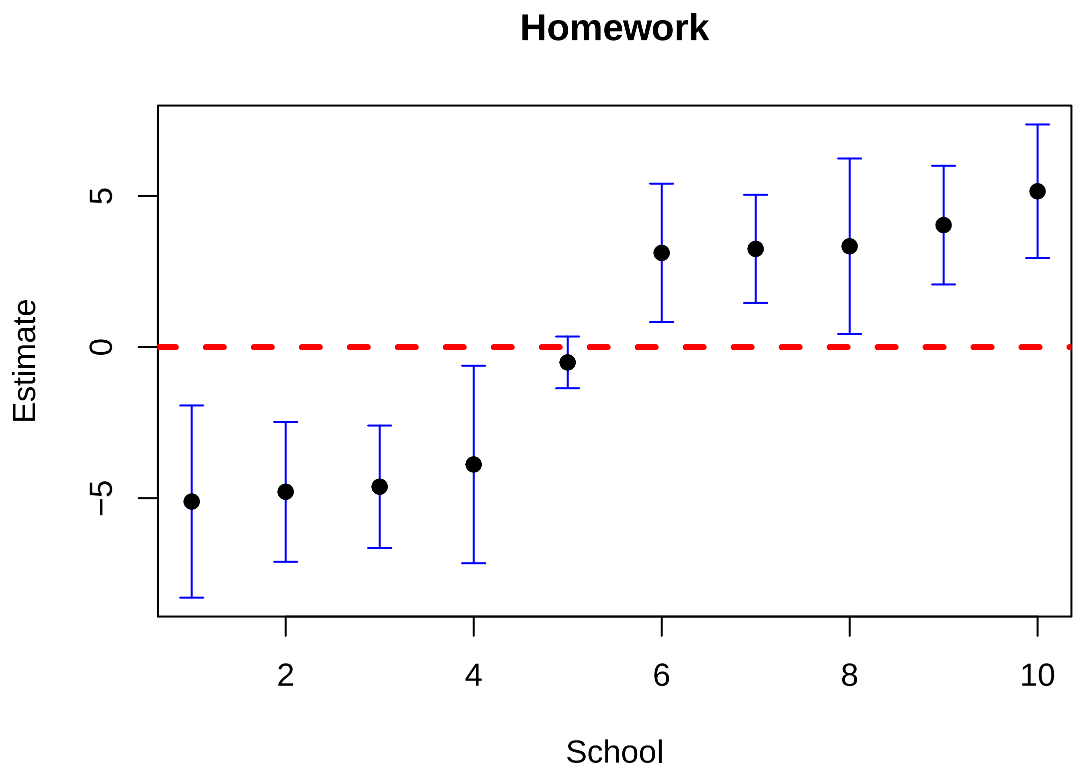
```
pander(res_hw[1:10,c("ID","y","upper","lower")],big.mark=",")
```

	ID	y	upper	lower
homework5	5	-5.11	-1.93	-8.289
homework1	1	-4.785	-2.47	-7.1
homework2	2	-4.618	-2.594	-6.643
homework6	6	-3.882	-0.6128	-7.152
homework7	7	-0.5051	0.352	-1.362
homework8	8	3.118	5.41	0.8263
homework4	4	3.251	5.04	1.461
homework9	9	3.338	6.243	0.4322
homework10	10	4.038	6.001	2.075
homework3	3	5.156	7.368	2.945

```
plotCI(1:nrow(res_int),res_int$y,ui=res_int$upper, li=res_int$lower,pch=19,scol="blue",xlab="School",yla
abline(h=mean(res_int$y),col=2,lwd=3,lty=2)
```




```
plotCI(1:nrow(res_hw),res_hw$y,ui=res_hw$upper, li=res_hw$lower,pch=19,scol="blue",xlab="School",ylab="
abline(h=mean(res_hw$y),col=2,lwd=3,lty=2)
```



Il modello interpreta bene i dati e tutti i parametri casuali sia quello relativo alle scuole che a "homework" è significativo. La correlazione tra il parametro relativo alle scuole e "homework" rimane negativo. "ses" come parametro fisso è significativo come "ratio" per un livello alpha di 0.05, mentre la parte fissa di "homework" non è significativa. Tra i parametri casuali inerenti le scuole il 6 e il 9 non sono significativi mentre per ciò che concerne homework non lo è il 7. Per entrambi i tipi di parametri casuali il modello postula graduatorie basati su valori attesi e intervalli di confidenza. Si aggiunge infine un modello random che ha fra le variabili esplicative anche l'interazione fra home e ratio.

```r
#-- R CODE
mod1 <- lmer(math ~ homework + ses + ratio + homework*ratio + (homework| schnum),d,REML=T)
summary(mod1)
```

Linear mixed model fit by REML ['lmerMod'] Formula: math ~ homework + ses + ratio + homework * ratio + (homework |
schnum) Data: d

REML criterion at convergence: 1741.7

Scaled residuals: Min 1Q Median 3Q Max -2.51889 -0.66423 0.00554 0.64440 2.69668

Random effects: Groups Name Variance Std.Dev. Corr schnum (Intercept) 45.62 6.754
homework 20.90 4.572 -0.90 Residual 41.30 6.427
Number of obs: 260, groups: schnum, 10

Fixed effects: Estimate Std. Error t value (Intercept) 60.8275 9.7098 6.265 homework -2.3550 6.5014 -0.362 ses 2.7144 0.6079 4.466 ratio -0.9516 0.6112 -1.557 homework:ratio 0.2701 0.4079 0.662

Correlation of Fixed Effects: (Intr) homwrk ses ratio homework -0.892

ses -0.028 -0.019
ratio -0.972 0.865 0.060
homework:rt 0.869 -0.973 0.011 -0.891

```
pander(Anova(mod1, type="III"),big.mark=",")
```

Table 8: Analysis of Deviance Table (Type III Wald chisquare tests)

	Chisq	Df	Pr(>Chisq)
(Intercept)	39.24	1	3.739e-10
homework	0.1312	1	0.7172
ses	19.94	1	7.985e-06
ratio	2.424	1	0.1195
homework:ratio	0.4386	1	0.5078

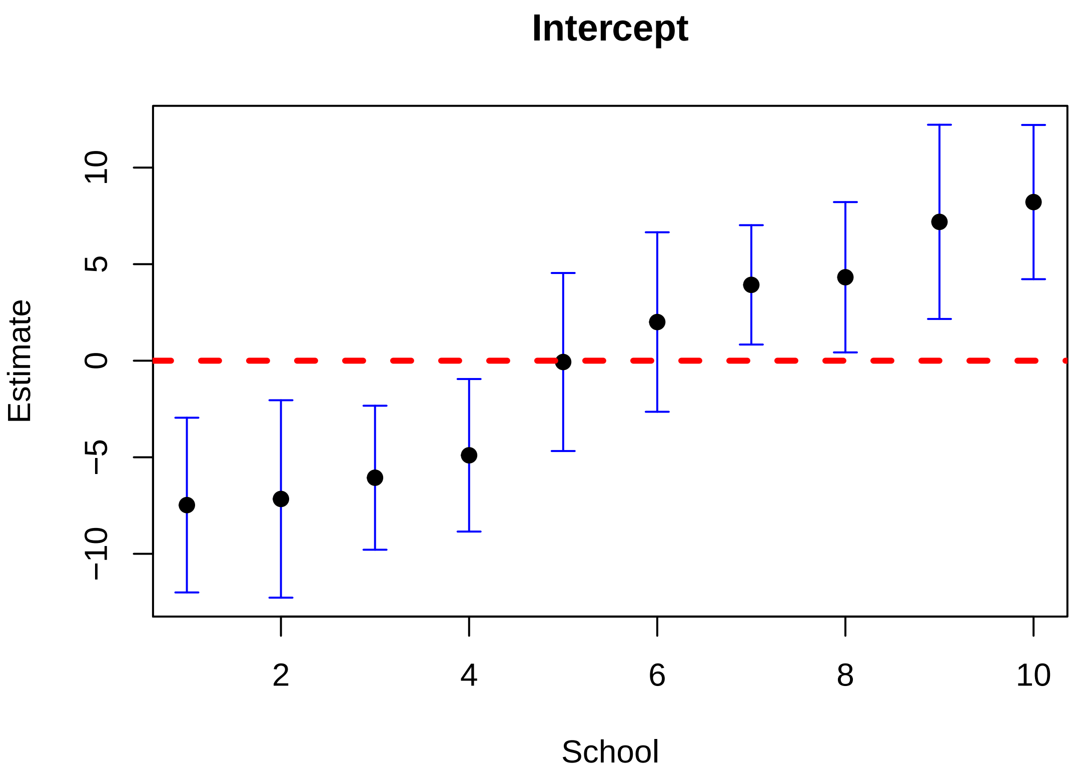
```
res <- sjp.lmer(mod1, type = "re.qq", sort.est = "sort.all",show.values=T,title="T",prnt.plot=F)
res$data$lower <- res$data$y-res$data$ci
res$data$upper <- res$data$y+res$data$ci

res_int <- subset(res$data,ind=="(Intercept)")
res_hw <- subset(res$data,ind=="homework")

pander(res_int[1:10,c("ID","y","upper","lower")],big.mark=",")
```

	ID	y	upper	lower
(Intercept)3	3	-7.476	-2.953	-12
(Intercept)8	8	-7.159	-2.047	-12.27
(Intercept)4	4	-6.06	-2.331	-9.788
(Intercept)10	10	-4.898	-0.9478	-8.848
(Intercept)9	9	-0.06695	4.543	-4.677
(Intercept)6	6	2.003	6.651	-2.644
(Intercept)7	7	3.928	7.019	0.8373
(Intercept)5	5	4.323	8.215	0.4301
(Intercept)2	2	7.19	12.22	2.161
(Intercept)1	1	8.215	12.21	4.222

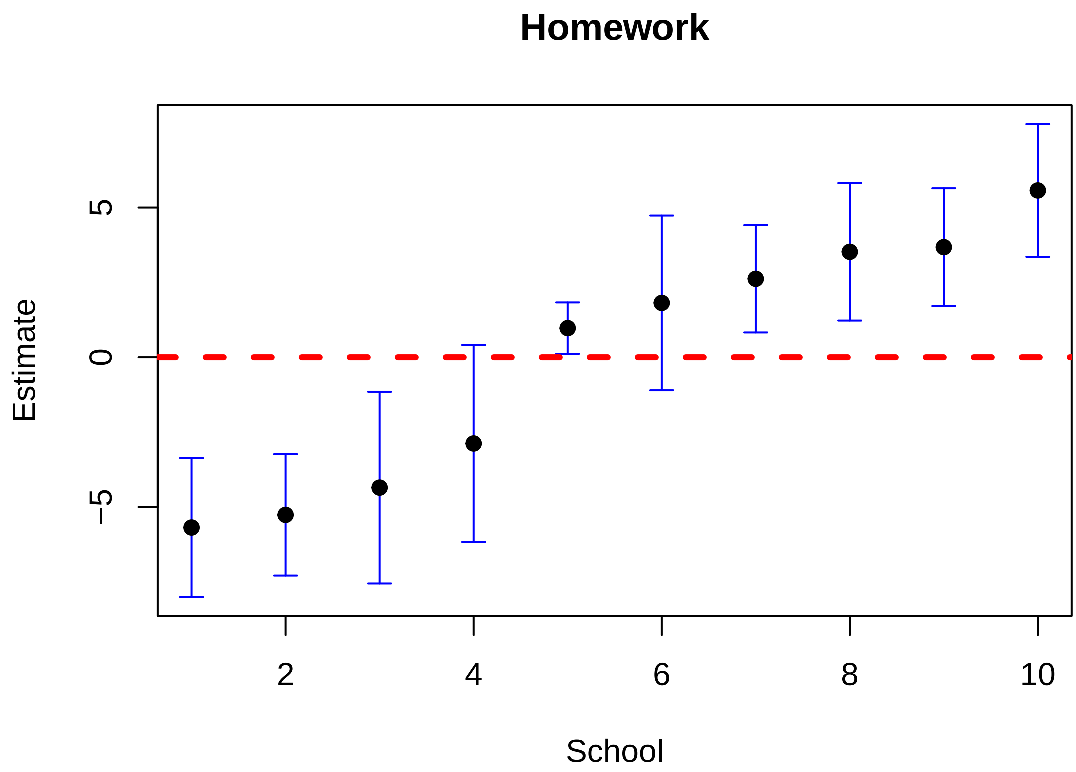
```
pander(res_hw[1:10,c("ID","y","upper","lower")],big.mark=",")
```

	ID	y	upper	lower
homework1	1	-5.685	-3.365	-8.006
homework2	2	-5.263	-3.236	-7.289
homework5	5	-4.352	-1.151	-7.554
homework6	6	-2.879	0.4105	-6.168
homework7	7	0.9743	1.832	0.117
homework9	9	1.815	4.733	-1.102
homework4	4	2.62	4.412	0.8286
homework8	8	3.522	5.817	1.226
homework10	10	3.677	5.643	1.711
homework3	3	5.571	7.786	3.355

```r
plotCI(1:nrow(res_int),res_int$y,ui=res_int$upper, li=res_int$lower,pch=19,scol="blue",xlab="School",yla
abline(h=mean(res_int$y),col=2,lwd=3,lty=2)
```




```r
plotCI(1:nrow(res_hw),res_hw$y,ui=res_hw$upper, li=res_hw$lower,pch=19,scol="blue",xlab="School",ylab="I
abline(h=mean(res_hw$y),col=2,lwd=3,lty=2)
```



Anche questo modello risulta significativo come tutti i parametri casuali inerenti scuole e “homework” e la loro correlazione che rimane negativa. In questo caso tra i parametri fissi però oltre a “homework” risultano non significativi anche ratio e l’interazione ratio-homework. Tra i parametri casuali inerenti le scuole il 5, 6, 7, 9, 10 non sono significativi mentre per ciò che concerne homework non lo è il 4, 6, 7, 9. Come si vede c’è un altro rilevante cambiamento dei ranking rispetto al modello precedente.