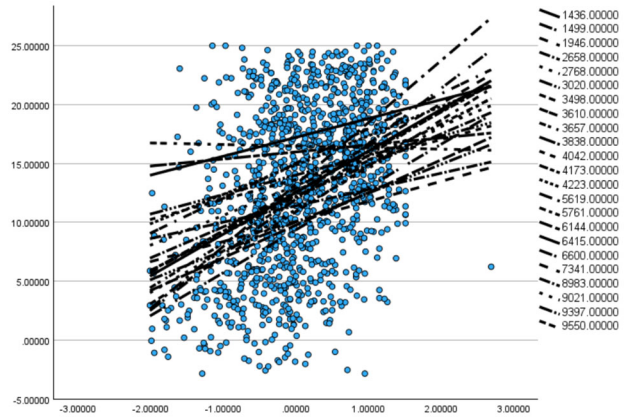
Plotting Within-Group Regression Lines: SPSS, R, and HLM
(For Hierarchically Structured Data)

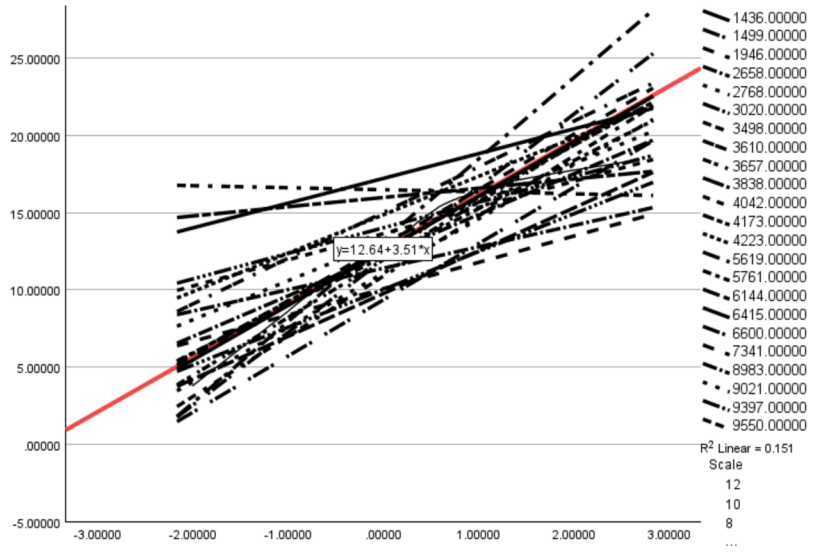
To illustrate plots of random slopes, I again used the random slopes model from the HSB data with SES as a predictor of math achievement. The results from that model indicated that the relationship (slope) between SES and math achievement varied significantly across schools. That means that SES was a better predictor of math achievement in some schools than in other schools. Plotting is essential for the researcher to help understand the results. For instance, from our results we don't really know whether SES always has a positive association with math achievement or whether in some schools the slope is flat or even negative. A plot of at least a sample of the slopes can be very helpful and, although certainly not required, is sometimes included in publications. Most studies have a large enough number of groups that plotting all of the slopes can be a bit too busy, so it is common to select a subsample. A random sample is usually recommended and there is no real rule about how many. I often feel 20 or so is sufficient to get an idea, but more may be desirable to get a better sense of the modal relationship.

SPSS

```
output close *.

get file='c:\jason\spsswin\mlrclass\hsbmerged.sav'.

* The CASESTOVARS command aggregates the data to select a random subset of schools.
* Note this step is needed because entire schools need to be selected  not just a subset of cases.
CASESTOVARS
 /ID=schoolid
 /DROP size to female .

* The SELECT IF command below could be used as an easy way to select the a subset of schools (commented).
* (This approach could be used before the casestovars command.)
* select if schoolid le 1637.

*A better method is to use a random sample.
* This sample command samples a .15 (15%)  proportion  groups from the full sample (N=160).
/* But you can also sample a certain number of cases, such as: SAMPLE 20 FROM 160.
sample .15.

*  The VARSTOCASES command disaggregates the data again to its original form of multiple records per school.
VARSTOCASES
 /MAKE ses from ses.1 TO ses.67
/MAKE mathach from mathach.1 TO mathach.67.

* The GGRAPH command draws an OLS linear regression line for each school and plots the slope.
GGRAPH
  /GRAPHDATASET NAME="GraphDataset" VARIABLES= ses mathach schoolid
/GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
SOURCE: s=userSource( id( "GraphDataset" ) )
DATA: ses=col( source(s), name( "ses" ) )
DATA: mathach=col( source(s), name( "mathach" ) )
DATA: schoolid = col(source(s), name("schoolid"), unit.category())
ELEMENT: point( position(ses * mathach))
ELEMENT: line(position(smooth.linear(ses * mathach)), shape(schoolid))
END GPL.
```



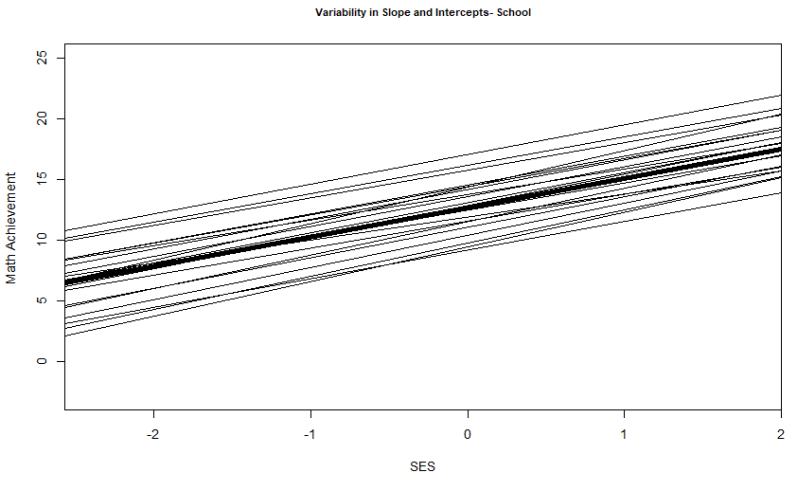
Clicking on the graph in the output and choosing Fit Line at Total under the Elements menu adds an average slope line, which can be edited to be a different color (choose Lines tab and change color) or thickness (choose Lines tab and adjust Weight). To remove the observed points, click on the graph and under the Options menu choose Un-Bin Element. Under the Marker tab choose white or the red diagonal line (no color) for the Fill and Border. I added the average fit line and changed the thickness in the plot below.



R

In the below example, I use a new package `lme4` and function `lmer` in R to make plotting easier. I omit the results except for the plot and just include the code below

```
library(lme4)
model <- lmer (mathach ~ ses +(ses|schoolid), data=mydata)
summary(model)

#sample 20 schools
ids <- sample(unique(mydata$schoolid), 20)
mydata2 <- mydata[mydata$schoolid %in% ids, ]

#names fixed  and random effects values from model and creates a combined data frame
fix <- fixef(model)
rand <- ranef(model)
paramsStudent <- cbind((rand$schoolid[1]+fix[1]),(rand$schoolid[2]+fix[2]))

#plot individual slopes for each school
plot(data = mydata2, mathach ~ ses,type = 'n',
     ylim = c(min(mydata2$mathach),max(mydata2$mathach)),
     xlim = c(min(mydata2$ses),max(mydata2$ses)),
     cex.main = .75,
     xlab = 'SES', ylab = "Math Achievement ",
     main = "Variability in Slope and Intercepts- School")

for(i in 1:length(unique(mydata2$schoolid))){
  abline(a = paramsStudent[i,1], b = paramsStudent[i,2],col = 'black')
  par<- par(new=F)
}

#add average slope (fixed effect)
abline(a = fix[1], b = fix[2], lwd= 5,col = 'black')
```



HLM

HLM graphs model results automatically. The built-in graphing functions are capable of a number of other kinds of graphs, but I demonstrate a basic one here (same general graph as obtained from SPSS)[1]. First, run the model with random slopes that was illustrated in the prior handout:

Level-1 Model

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

Level-2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

SES has been centered around the grand mean.

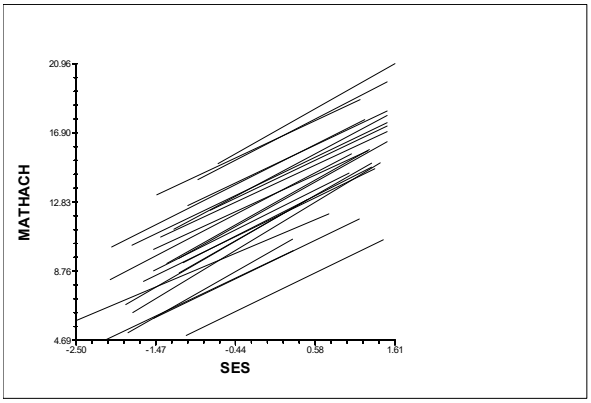
Mixed Model

$$MATHACH_{ij} = \gamma_{00} + \gamma_{10}*SES_{ij} + u_{0j} + u_{1j}*SES_{ij} + r_{ij}$$

Then use the menus to obtain the graph.

File → Graph Equations → Level-1 equation graphing.

Under "X-focus", choose the Level-1 predictor (here, it was SES). Under **Number of groups** choose whether you would to graph a random sample of groups, the first 10 groups, or all groups. If you choose a random sample (as I did here), input the probability (e.g., .15 for 15% of groups). Click **Ok.**



[1] Note that the R and HLM methods use a "model" graph which creates graphs based on the results of the model. So the lines are based on empirical Bayes estimates of the slopes rather than OLS. SPSS is producing OLS estimates which may appear to vary more across groups. The three graphs also differ because a different sample of groups was drawn in the three graphs and default choices for axis dimensions. I will discuss the empirical Bayes concept and the difference from OLS soon.