Summer 2013 –Multilevel Models for Applied Cross-Sectional Data Workshop

Lecture #1 Annotated SAS Handout

The following handout accompanies the lecture notes for Section 1. In the lecture, a simulated data set was used to illustrate regression and ANOVA under the context of the General Linear Model and to show how hierarchical analyses can show differential patters of results at different levels.

Simulation Code:

```
*Code to generate example data;
data school1 (keep=student school studentses achieve);

*sets the random seed so everyone's data is the same;
      seed=14;


*simulate seven schools;
      do school=1 to 7;

*random intercept for school has a variance of 16;
            schoolint=rannor(seed)*sqrt(16);
*school mean SES - grand mean of 50, variance of 9;
            schoolses=3*rannor(seed)+50;

*simulate 50 students within each school;
            do student=1 to 50;

*error variance = 5^2 = 25;
                  error=sqrt(25)*rannor(seed);

*SES is on a 0-100 metric with mean 50 and SD of 2;
                  studentses = sqrt(4)*rannor(seed)+schoolses;
*linear model - studentses is group mean centered.
*Overall intercept is 0;
*Slope for school mean SES is 2 - between school effect
*Slope for student mean-centered SES is -1 - within school effect;
*achieve is the dependent variable;
                  achieve = (schoolint+2*schoolses)+-1*(studentses-schoolses)+error;
                  output;
            end;
      end;
run;
```

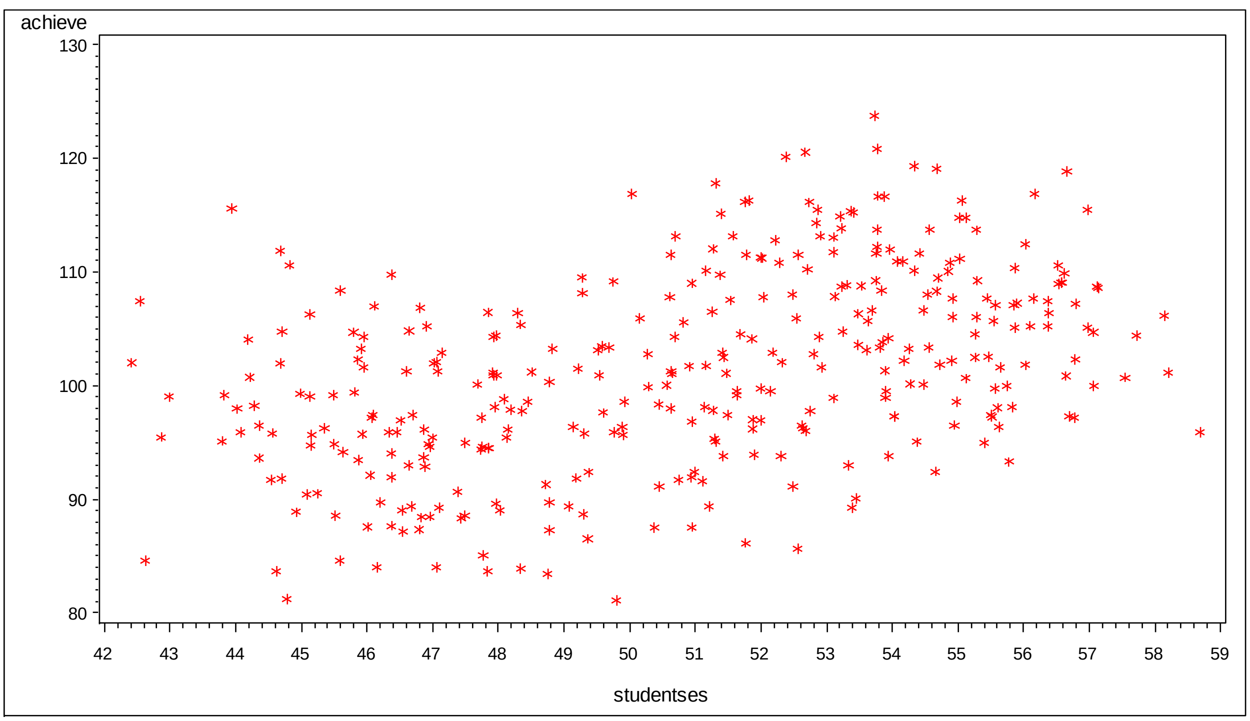
```
*export simulated data to a csv file for analysis - change to your path;
proc export data=school1
            outfile= "C:\data.csv"
            dbms=csv label replace;
            putnames=yes;
run;
```

Analysis Code and Output:

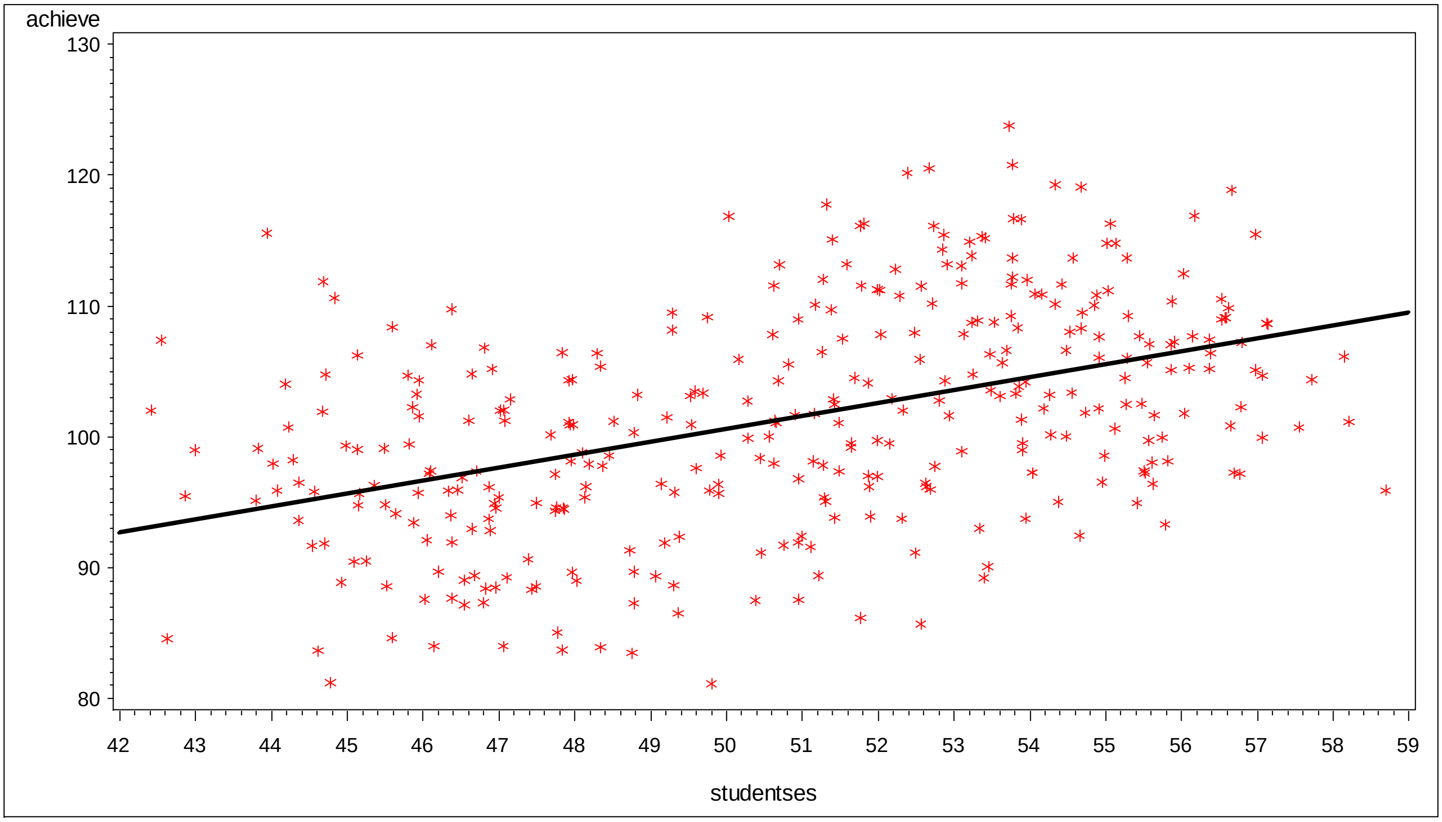
```
*code to import the data into SAS - change to your path;
proc import out=school1
            datafile= "C:\schooldata.csv"
            dbms=csv replace;
            getnames=yes;
            datarow=2;
run;
```

```
*plot acheivement by student ses for all data;
*reset gplot options;
goptions reset=all border;
proc gplot data=school1;
*select stars as symbols;
symbol1 value=star cv=red ci=black co=blue width=4;
*make the first plot - no regression line;
plot achieve*studentses; run;
*ask for an overall regression line;
symbol1 interpol=rl value=star cv=red ci=black co=blue width=4;
*make the second plot - with overall regression line;
plot achieve*studentses; run;
quit;
```

Resulting Graph #1 – data without regression line



Resulting Graph #2 – Data with regression line:



Data Analysis

```
*preliminary analysis: descriptive statistics;
proc means data=school1;
var achieve studentses;
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
achieve	350	101.3811912	8.5143419	81.1617979	123.7855178
studentses	350	50.8501476	3.8808480	42.4146209	58.6925801

```
*Analysis #1 - empty regression model with just the intercept - using proc reg;
proc reg data=school1;
model achieve = ;
run;
```

```
                          The REG Procedure
                            Model: MODEL1
                     Dependent Variable: achieve

                Number of Observations Read          350
                Number of Observations Used          350


                         Analysis of Variance

                                      Sum of           Mean
Source                   DF          Squares         Square    F Value    Pr > F

Model                     0                0              .          .
Error                   349            25300       72.49402
Corrected Total         349            25300

             Root MSE              8.51434    R-Square     0.0000
             Dependent Mean      101.38119    Adj R-Sq     0.0000
             Coeff Var             8.39834


                         Parameter Estimates

                              Parameter       Standard
        Variable      DF       Estimate          Error    t Value    Pr > |t|

        Intercept      1      101.38119        0.45511     222.76      <.0001
```

$\sigma^2 = 79.49$

$\sigma = 8.51 = SD(Y)$

$\beta_0 = 101.38 = \bar{Y}$

```
*Analysis #2 - regression model predicting student acheivement by student SES - in proc
mixed;
proc mixed data=school1;
model achieve=studentses/s;
run;
```

```
                 Covariance Parameter
                       Estimates

                 Cov Parm      Estimate

                 Residual       58.0604


                    Fit Statistics

         -2 Res Log Likelihood          2415.4
         AIC (smaller is better)        2417.4
         AICC (smaller is better)       2417.4
         BIC (smaller is better)        2421.3


              Solution for Fixed Effects

                             Standard
Effect         Estimate         Error       DF    t Value     Pr > |t|

Intercept       51.3152        5.3598      348       9.57       <.0001
studentses       0.9846        0.1051      348       9.37       <.0001

             Type 3 Tests of Fixed Effects

                        Num      Den
         Effect          DF       DF     F Value     Pr > F

         studentses       1      348       87.76     <.0001
```

$\sigma^2 = 58.06$

$\beta_0 = 51.32$

$\beta_1 = 0.98$

β_1 is significant (p <0.0001)

```
*Analysis #3 - ANOVA model for differences in acheivement;
proc mixed data=school1;
class school;
model achieve=school/s;
lsmeans school;
run;
```

```
                    Covariance Parameter
                         Estimates

                    Cov Parm     Estimate

                    Residual      29.0378


                       Fit Statistics

              -2 Res Log Likelihood         2156.2
              AIC (smaller is better)       2158.2
              AICC (smaller is better)      2158.2
              BIC (smaller is better)       2162.0
```

$\sigma^2 = 29.04$

The Mixed Procedure

Solution for Fixed Effects

Effect	school	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		93.0464	0.7621	343	122.10	<.0001
school	1	-0.5444	1.0777	343	-0.51	0.6138
school	2	15.3386	1.0777	343	14.23	<.0001
school	3	6.5288	1.0777	343	6.06	<.0001
school	4	8.4666	1.0777	343	7.86	<.0001
school	5	18.4930	1.0777	343	17.16	<.0001
school	6	10.0608	1.0777	343	9.34	<.0001
school	7	0	.	.	.	.

$\beta_0 = 93.05$

Intercept is the mean of the reference group (school 7)

Regression coefficients represent mean differences of schools from reference group (school 7)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
school	6	343	88.05	<.0001

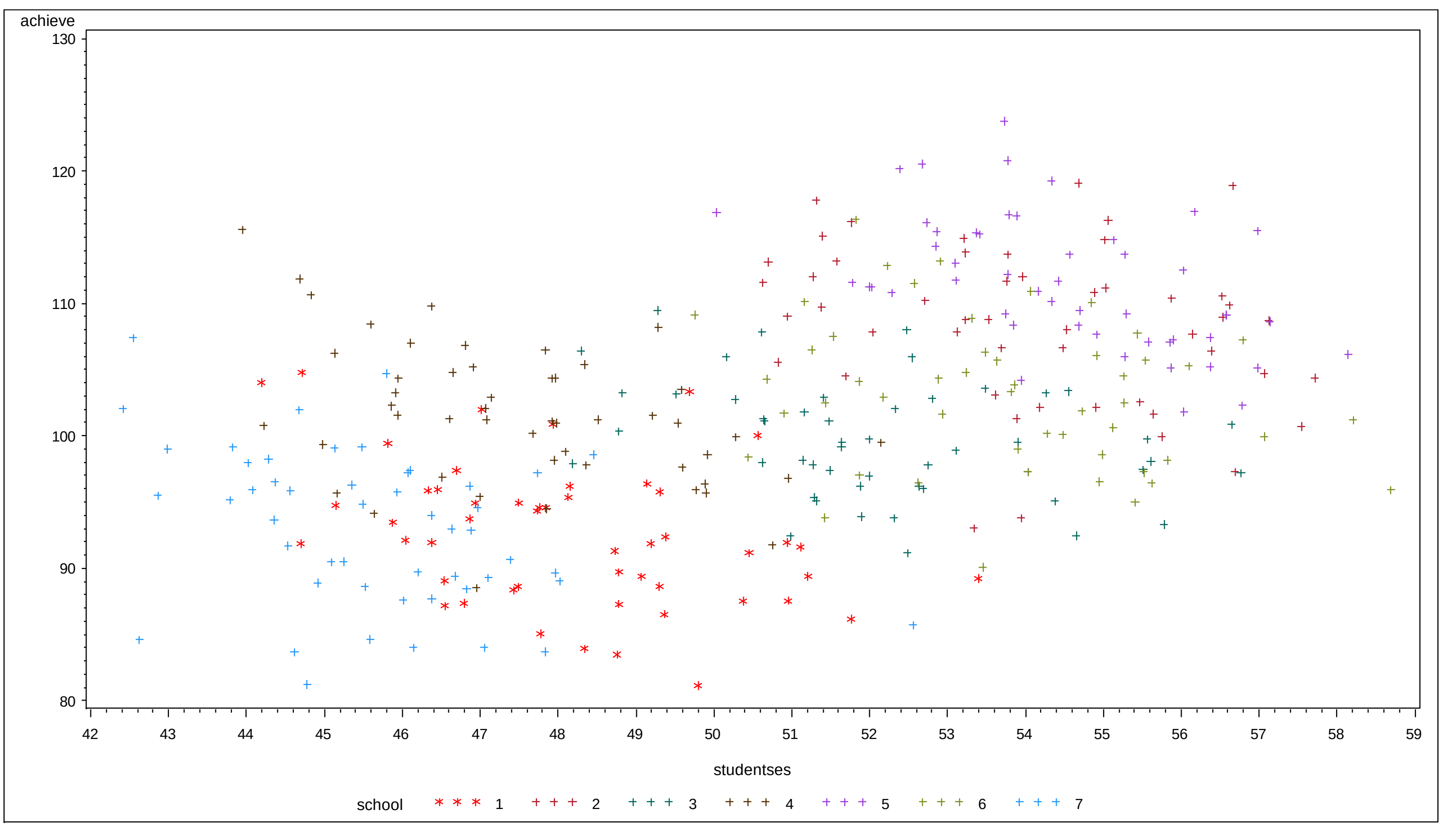
There is a significant difference between at least one pair of school means (p < 0.001)

```
                         Least Squares Means

                                       Standard
Effect     school         Estimate       Error      DF    t Value    Pr > |t|

school             1       92.5020      0.7621     343     121.38      <.0001
school             2        108.39      0.7621     343     142.22      <.0001
school             3       99.5752      0.7621     343     130.66      <.0001
school             4        101.51      0.7621     343     133.21      <.0001
school             5        111.54      0.7621     343     146.36      <.0001
school             6        103.11      0.7621     343     135.30      <.0001
school             7       93.0464      0.7621     343     122.10      <.0001
```

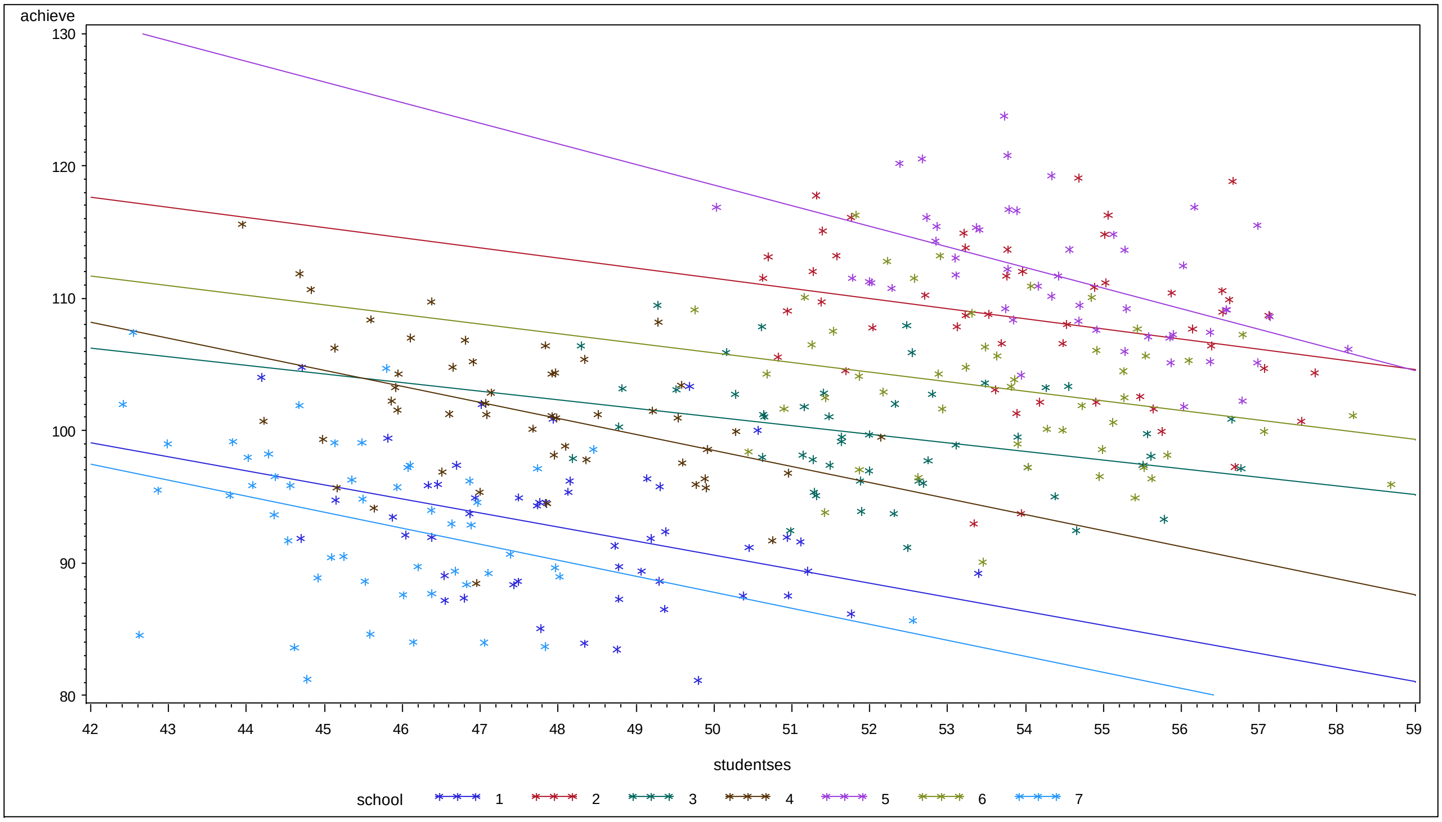
List of school means for comparison with regression weights

```
*plot acheivement by student ses - by school;
goptions reset=all border;
proc gplot data=school1;
*make the first plot - no regression line;
plot achieve*studentses=school; run;
*make the second plot - with overall regression line;
symbol1 interpol=rl value=star width=1 repeat=1000;
plot achieve*studentses=school; run;
quit;
```

Plot #1 – colors indicate school



Plot #2 – each school has its own regression line – indicating a different intercept per school.



Get the school mean SES for each school to incorporate into analysis:

```
*first sort the data by school (SAS requirement);
proc sort data=school1;
by school;
run;

*second get the mean studentSES and put into a new data set called school_means;
proc means data=school1;
by school;
var studentses;
output out=school_means mean(studentses)=schoolmean;
run;

*third - open the school_means data set and keep only the relevant variables;
data school_means (keep=school schoolmean);
set school_means;
run;

*fourth - merge the school_means data set with the whole data set and subtract
*the school mean from each student's SES, creating a cluster mean centered variable;
data school2 (keep=school student achieve studentses schoolmean studentsesM);
merge school1 school_means;
by school;
studentsesM=studentses-schoolmean;
run;
```

```
*Analysis #4 - Adding school level;
proc mixed data=school2 covtest;
class school;
model achieve=schoolmean studentses/s ddfm=bw;
random int/subject=school g;
run;
```

```
                              The Mixed Procedure

                               Estimated G Matrix

                 Row     Effect        school              Col1

                   1     Intercept                1      18.8901


                         Covariance Parameter Estimates

                                               Standard         Z
          Cov Parm       Subject     Estimate       Error     Value       Pr > Z

          Intercept      school       18.8901     12.2688      1.54       0.0618
          Residual                    25.4289      1.9446     13.08       <.0001


                                Fit Statistics

                    -2 Res Log Likelihood          2147.1
                    AIC (smaller is better)        2151.1
                    AICC (smaller is better)       2151.1
                    BIC (smaller is better)        2151.0


                           Solution for Fixed Effects

                                        Standard
           Effect          Estimate        Error       DF     t Value     Pr > |t|

           Intercept        18.6914      25.1981        5        0.74       0.4916
           schoolmean        2.6161       0.5140        5        5.09       0.0038
           studentses       -0.9900       0.1405      342       -7.05       <.0001


                          Type 3 Tests of Fixed Effects

                                 Num      Den
                  Effect          DF       DF     F Value     Pr > F

                  schoolmean       1        5       25.90     0.0038
                  studentses       1      342       49.68     <.0001
```

$\tau_0^2 = 29.04$

$\sigma^2 = 25.43$

$\gamma_{00} = 18.69$ (overall intercept)

$\gamma_{01} = 2.61$ (level 2) (school mean slope)

$\gamma_{10} = -0.99$ (level 1) (student slope)