Correlated Data

Statistical Inference II

April 4, 2018

General Linear Models require responses to be approximately normally distributed and independent. Through generalized linear models, we learned about handling responses that are not normally distributed (e.g. Poisson, Binomial). From now on we learn about multilevel models / linear mixed effect models / hierarchical linear models that can model response that violate the independence assumption. These models have responses that are correlated thus not independent.

e.g.

In this class and in the next few, we will be using the `hsb.csv` dataset on Canvas>Modules>Datasets. We will try to understand math achievement (`mathach`) of students based on their socio-economic status (`ses`). We will use the notation Y_{ij} for math achievement of ith student in the jth school. There are _____ students within _____ schools in this dataset.

Using R, set random seed to your phone number without the area code.

Calculate the average math achievement score for all students in the dataset.

$\bar{Y}_{..} =$

Pick a random school (make sure each school has equal probability to be selected). For **only** the school you selected, find the mean math achievement.

$\bar{Y}.j =$

For your school, fit a general linear model where math achievement is the response and ses is the predictor. Record the coefficients:

$\hat{\beta}_{0j}$: $\hat{\beta}_{1j}$:

For your school, fit a general linear model where grand mean cenetered math achievement is the response and ses is the predictor. Record the coefficients:

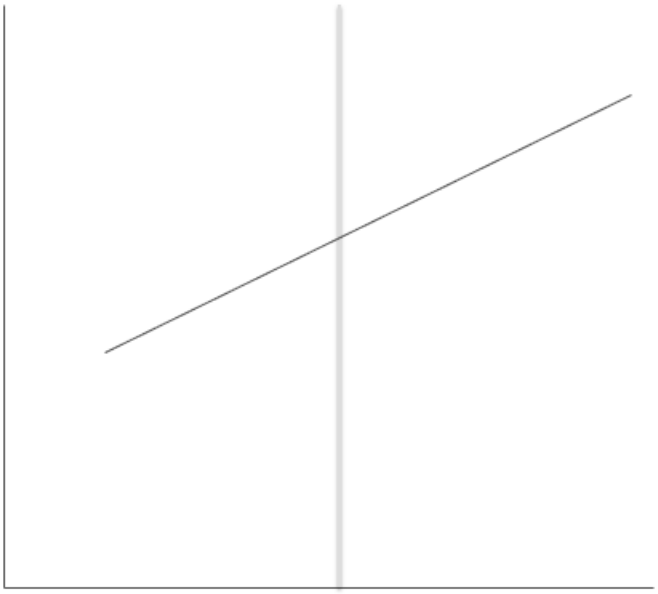
$\hat{\beta}_{0j}$: $\hat{\beta}_{1j}$:

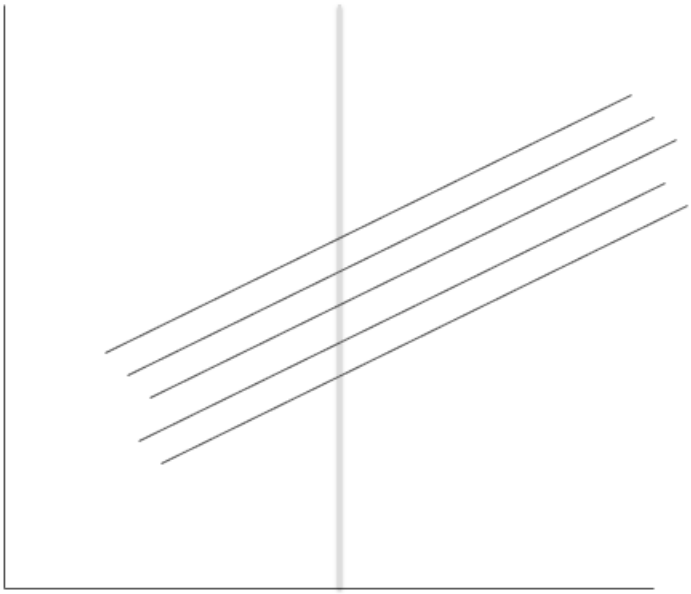
For your school, fit a general linear model where group mean cenetered math achievement is the response and ses is the predictor. Record the coefficients:

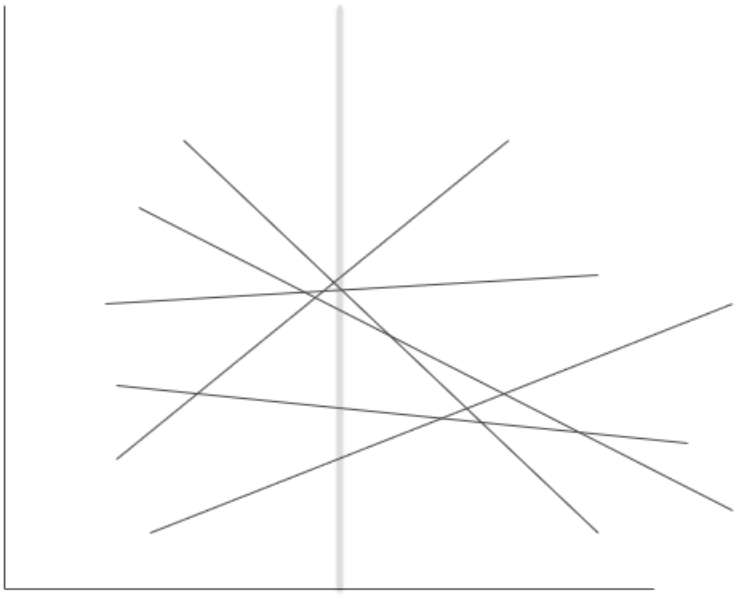
$\hat{\beta}_{0j}$: $\hat{\beta}_{1j}$:

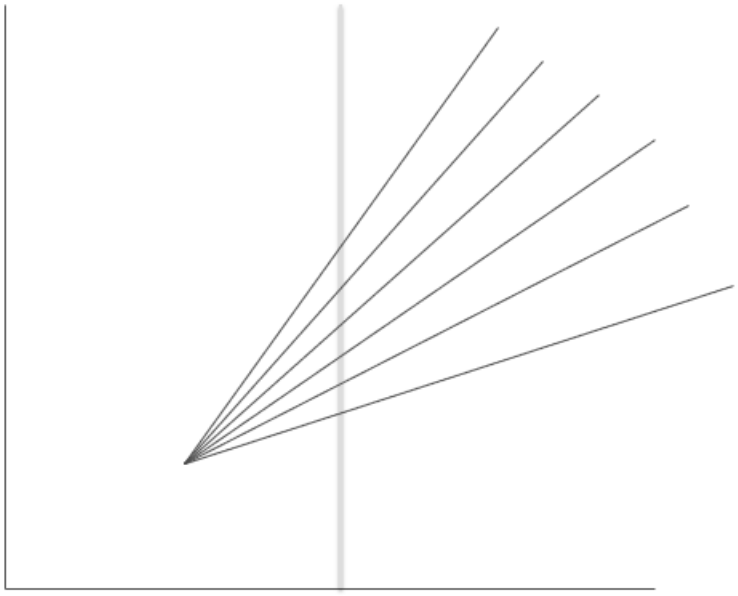
We will use group-mean centering moving on.

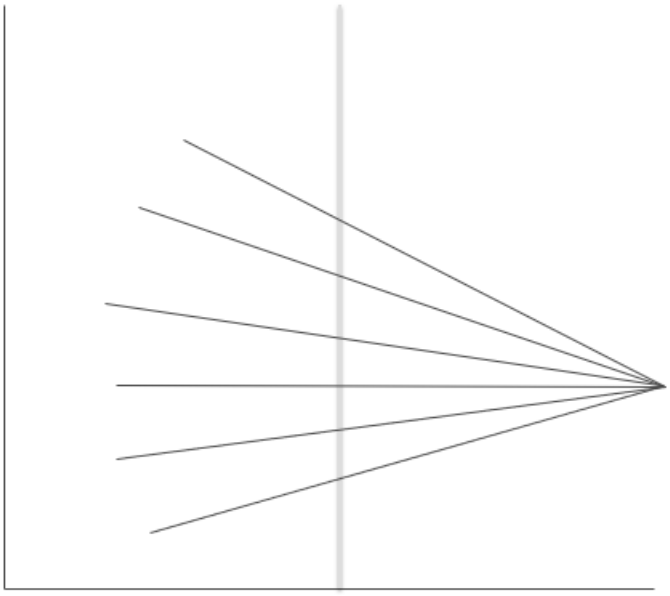
Different effect types











Empty Model - One-Way Random-Effect ANOVA model

```
library(lme4)
model1<-lmer(mathach~1+(1|schoolid),data=hsb)
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathach ~ 1 + (1 | schoolid)
##    Data: hsb
##
## REML criterion at convergence: 47116.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0631 -0.7539  0.0267  0.7606  2.7426
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  8.614   2.935
##  Residual             39.148   6.257
## Number of obs: 7185, groups:  schoolid, 160
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  12.6370     0.2444   51.71
```

Model Notation:

Parameter Estimates:

ICC