COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

# STAT 4224/5224

***Bayesian Statistics***

Dobrin Marchev

# Introduction

- We learned about the concept of hierarchical modeling, a data analysis approach that is appropriate when we have multiple measurements within each of several groups.
- Now we extend the hierarchical model to describe how relationships between variables may differ between groups. \
- This can be done with a regression model to describe within-group variation, and a multivariate normal model to describe heterogeneity among regression coefficients across the groups.
- We also cover estimation for hierarchical generalized linear models, which are hierarchical models that have a generalized linear regression model representing within-group heterogeneity.

# Hierarchical Regression

Let's return to the math score data described in Chapter 8, which included math scores of 10th grade children from 100 different large urban public high schools. Back then we estimated school-specific expected math scores, as well as how these expected values varied from school to school.

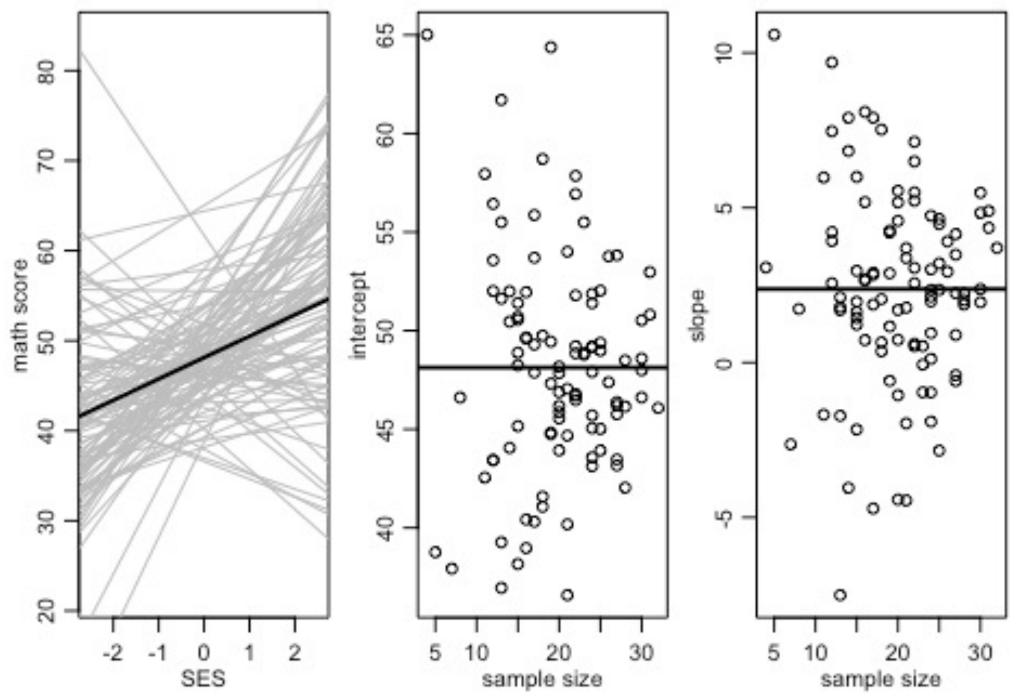
Now suppose we are interested in examining the relationship between math score and another variable, socioeconomic status (SES), which was calculated from parental income and education levels for each student in the dataset.

Given the amount of variation we observed it seems possible that the relationship between math score and SES might vary from school to school as well.

A quick and easy way to assess this possibility is to fit a linear regression model of math score as a function of SES for each of the 100 schools in the dataset.

# Results per school

The second and third panels of the figure relate the least squares estimates to sample size. Notice that schools with the highest sample sizes have regression coefficients that are generally close to the average, whereas schools with extreme coefficients are generally those with low sample sizes.



# Hierarchical Model

We use an ordinary regression model to describe within-group heterogeneity of observations, then describe between-group heterogeneity using a sampling model for the group-specific regression parameters.

$$Y_{ij} = \boldsymbol{\beta}'_j \boldsymbol{x}_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim iid\ N(0, \sigma^2)$$

where $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of regressor for observation $i$ in group $j$.

Alternatively,

$$\boldsymbol{Y}_j \sim N\left(\boldsymbol{X}_j \boldsymbol{\beta}_j, \sigma^2 \boldsymbol{I}\right), j = 1, \dots, m$$

he heterogeneity among the regression coefficients $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$ will be described with a between-group sampling model:

$$\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m \sim iid\ N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$$

# Mixed Effects Model

This hierarchical regression model is sometimes called a linear mixed effects model. It can be rewritten as

$$\beta_j = \boldsymbol{\theta} + \boldsymbol{\gamma_j}$$
$$\gamma_1, \dots, \gamma_m \sim iid\ N(0, \Sigma)$$

This means:

$$Y_{ij} = \boldsymbol{\beta}'_j \boldsymbol{x}_{ij} + \varepsilon_{ij} = \boldsymbol{\theta}' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}'_j \boldsymbol{x}_{ij} + \varepsilon_{ij}$$

In this parameterization $\boldsymbol{\theta}$ is referred to as a fixed effect as it is constant across groups, whereas $\gamma_1, \dots, \gamma_m$ are called random effects as they vary between the groups. The name "mixed effects model" comes from the fact that the regression model contains both fixed and random effects. The regressors can be different for the different effects:

$$Y_{ij} = \boldsymbol{\theta}' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}'_j \boldsymbol{z}_{ij} + \varepsilon_{ij}$$

In particular, $x_{ij}$ might contain regressors that are group specific, that is, constant across all observations in the same group.

# Bayesian Analysis

The goal is to estimate the full posterior

$$f(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m, \theta, \Sigma, \sigma^2 | X_1, \dots, X_m, y_1, \dots, y_m)$$

It can still be achieved with Gibbs sampler if we assume semi conjugate priors:

$$\theta \sim N(\boldsymbol{\mu}_o, \boldsymbol{\Lambda}_0)$$

$$\Sigma \sim IW(\eta_0, S_0)$$

$$\sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

# Conditional Distributions

Conditional on $\theta, \Sigma, \sigma^2$ the regression coefficients $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$ are independent, each having a multivariate normal distribution with mean, and covariance as follows:

$$cov(\boldsymbol{\beta}_j | y_j, X_j, \sigma^2, \theta, \Sigma) = \left(\Sigma^{-1} + \frac{X_j' X_j}{\sigma^2}\right)^{-1}$$

$$E(\boldsymbol{\beta}_j | y_j, X_j, \sigma^2, \theta, \Sigma) = \left(\Sigma^{-1} + \frac{X_j' X_j}{\sigma^2}\right)^{-1} \left(\Sigma^{-1}\theta + \frac{X_j' y_j}{\sigma^2}\right)$$

The other full conditionals are also possible to be derived:

$$\theta | \beta_1, \dots, \beta_m, \Sigma \sim N$$

$$\Sigma | \theta, \beta_1, \dots, \beta_m, X_1, \dots, X_m, y_1, \dots, y_m \sim IG$$

# Example (p. 200)

For the school data example, we must choose priors:

$$\theta \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$$

This is the overall school level regression parameter vector.

We choose $\boldsymbol{\mu}_0$ be equal to the average of the ordinary least squares regression estimates of all schools.

$\boldsymbol{\Lambda}_0$ is chosen to be the sample covariance of all beta vectors of all OLS.

$$\Sigma \sim IW(\eta_0, \boldsymbol{S}_0)$$

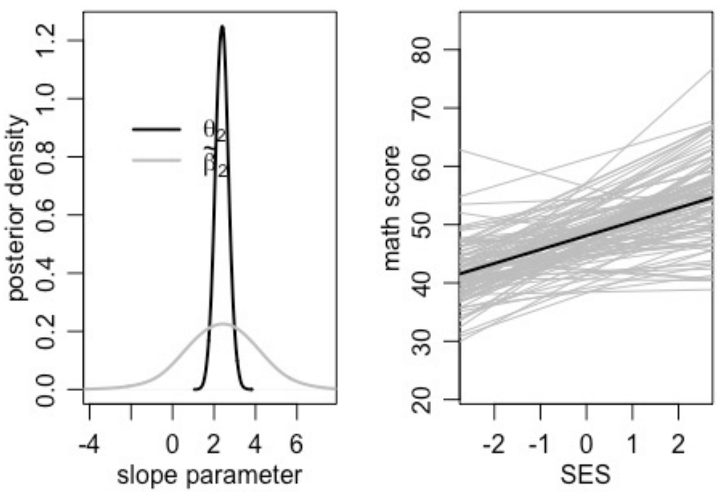
We will take $\boldsymbol{S}_0$ to be the same as $\boldsymbol{\Lambda}_0$

For $\eta_0$ we choose $p + 2 = 4$ making it diffuse.

$$\sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

$$\nu_0 = 1$$

# Example (p. 200)




The first panel plots the posterior density of the expected slope $\theta_2$ of a randomly sampled school, as well as the posterior predictive distribution of a randomly sampled slope. The second panel gives posterior expectations of the 100 school-specific regression lines, with the average line given in black.