ADVANCED BAYESIAN MODELING

Grouping and Regression Coefficients

Observations are divided into J groups, with scalar responses

$$y_{ij} \quad = \quad \text{response from obs. } i \text{ in group } j \qquad i = 1, \ldots, n_j \quad j = 1, \ldots, J$$

The regression has K explanatory variables (perhaps including an intercept). For observation i in group j,

$$X_{ij} \; = \; (x_{ij,1}, \ldots, x_{ij,K})$$

Note: The explanatory variables must *not* include indicators for the J groups!

Ordinarily, linear regression would be

$$\mathrm{E}(y_{ij} \mid \theta, X_{ij}) \;=\; X_{ij}\beta \;=\; \beta_1 x_{ij,1} + \cdots + \beta_K x_{ij,K}$$

Now instead allow coefficients to vary by group:

$$\mathrm{E}(y_{ij} \mid \theta, X_{ij}) \;=\; X_{ij}\beta^{(j)} \;=\; \beta_1^{(j)} x_{ij,1} + \cdots + \beta_K^{(j)} x_{ij,K}$$

$$\text{where} \qquad \beta^{(j)} \;=\; \begin{pmatrix} \beta_1^{(j)} \\ \vdots \\ \beta_K^{(j)} \end{pmatrix}$$

Example: Rat Growth Data

$$y_{ij} \quad = \quad \text{weight of rat } j \text{ at age } x_i \qquad i = 1, \ldots, 5 \quad j = 1, \ldots, 30$$

Ages happened to be same for all rats:

$x_1 = 8$ days $\quad x_2 = 15$ days $\quad x_3 = 22$ days $\quad x_4 = 29$ days $\quad x_5 = 36$ days

(but this is not necessary for the methods to work)

Data Source: Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**, 972–985.

At early ages, growth of a rat is well modeled by a straight line:

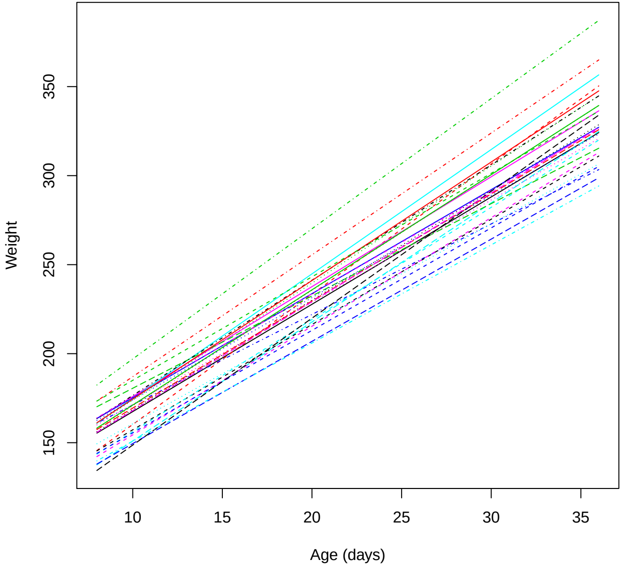
$$\text{weight} \quad \approx \quad \beta_1 \; + \; \beta_2 \times \text{age}$$

But different rats could have

- Different intercepts β_1: Some weigh more than others at birth.
- Different slopes β_2: Some grow faster than others.

Individual ordinary least squares fitted simple linear regression lines (weight vs. age) for the 30 rats

Note: Some rats consistently weigh more, and some grow faster.



Thus, consider regression with coefficients allowed to vary by rat (group):

$$\mathrm{E}(y_{ij} \mid \theta, X_{ij}) \;=\; \beta_1^{(j)} + \beta_2^{(j)} x_i$$

where

$$y_{ij} \;=\; \text{weight of rat } j \text{ at age } x_i$$

Note: In earlier notation,

$$x_{ij,1} \;=\; 1 \qquad\qquad x_{ij,2} \;=\; x_i$$