SAT Coaching Study example

- Example from Gelman et.al, section 5.5
- Separate randomized experiments in 8 high schools
- Treatment is local SAT coaching program
- Outcome is SAT-Verbal score (200 to 800)
- Treatment effect estimated using analysis of covariance
- Need to estimate the average effects of coaching program for each school.

Data:

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j	Average treatment effect, θ_j
A	28	15	?
B	8	10	?
C	−3	16	?
D	7	11	?
E	−1	9	?
F	1	11	?
G	18	10	?
H	12	18	?

Preliminary data analysis:

Two approaches:

1. Separate estimates - Consider the 8 programs separately
   - two programs appear to work (18-28 points)
   - four programs appear to have a small effect
   - two programs appear to have negative effects

2. Pooled estimate – Combine all schools into single analysis. We'll then obtain a pooled estimate of about 8 with standard error 4.2

Neither separate nor pooled estimates seem right.

Hierarchical model provides a compromise between the two estimates

Hierarchical Model -

- Data in each school is modeled as depending on “true school effect”
- Eight “true school effects” are thought of as coming from a population of possible school effects (described by a Gaussian distribution)
- Bayesian computation finds posterior distribution of true school effects and summarizes the population distribution of school effects
- Hierarchical models are applicable where data spans multiple clusters and each has cluster-specific parameters. Theses parameters are thought to be deviations from the overall mean mu.

- Data model / likelihood:

  $$y_j|\theta_j \overset{\text{ind}}{\sim} \mathrm{N}(\theta_j, \sigma_j^2), \text{ for } j = 1, \ldots, 8$$

  with σ_j^2's assumed known

  - normality and known variance justified by large sample size in each school

- Prior distribution: $\theta_j|\mu, \tau^2 \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \tau^2)$ for $j = 1, \ldots, 8$
  - exchangeable prior distn for θ_j's
  - traditional random effects model
    - ∗ note $\tau \to 0$ reduces to complete pooling
    - ∗ note $\tau \to \infty$ reduces to separate estimates

The joint posterior distribution we are trying to estimate is given below

- Joint posterior distribution:

$$
\begin{aligned}
& p(\theta, \mu, \tau | y) \\
& \quad \propto p(y|\theta) p(\theta|\mu, \tau) p(\mu, \tau) \\
& \quad \propto \prod_{j=1}^{8} N(y_j|\theta_j, \sigma_j^2) \prod_{j=1}^{8} N(\theta_j|\mu, \tau^2) \\
& \quad \propto \tau^{-8} \exp\left[-\frac{1}{2}\sum_j \frac{1}{\tau^2}(\theta_j - \mu)^2\right] \exp\left[-\frac{1}{2}\sum_j \frac{1}{\sigma_j^2}(y_j - \theta_j)^2\right]
\end{aligned}
$$

Computation of this posterior distribution using Gibbs sampling -

- To simulate from joint posterior distribution $p(\theta, \mu, \tau|y)$:
  1. draw τ from $p(\tau|y)$ (grid approximation)
  2. draw μ from $p(\mu|\tau, y)$ (normal distribution)
  3. draw $\theta = (\theta_1, \ldots, \theta_J)$ from $p(\theta|\tau, y)$
     (independent normal distribution for each θ_j)

Results -

School	Posterior quantiles					Estimates	
	2.5%	25%	50%	75%	97.5%	pooled	separate
A	−2	6	10	16	32	8	28
B	−5	4	8	12	20	8	8
C	−12	3	7	11	22	8	−3
D	−6	4	8	12	21	8	7
E	−10	2	6	10	17	8	−1
F	−9	2	6	10	19	8	1
G	−1	6	10	15	27	8	18
H	−7	4	8	13	23	8	12
μ	−2	5	8	11	18		
τ	0.3	2.3	5.1	8.8	21.0		

As we can infer from the results, Hierarchical model provides better estimates than pooled and separate analysis.