1 Hierarchical models: data-analysis problems

1.1 Math tests

The data set in "mathtest.csv" shows the scores on a standardized math test from a sample of 10th-grade students at 100 different U.S. urban high schools, all having enrollment of at least 400 10th-grade students. (A lot of educational research involves "survey tests" of this sort, with tests administered to all students being the rare exception.)

Let θ_i be the underlying mean test score for school i, and let y_{ij} be the score for the jth student in school i. Starting with the "mathtest.R" script, you'll notice that the extreme school-level averages $\bar{y}_i$ (both high and low) tend to be at schools where fewer students were sampled.

1. Explain briefly why this would be.

   In schools where few children are sampled, if one student does extremely well or extremely poorly, the average is heavily influenced by that score. The more students are sampled, the less any one score can skew the average.

2. Fit a normal hierarchical model to these data via Gibbs sampling:

$$\begin{aligned} y_{ij} &\sim \mathrm{N}(\theta_i, \sigma^2) \\ \theta_i &\sim \mathrm{N}(\mu, \tau^2\sigma^2) \end{aligned}$$

   Decide upon sensible priors for the unknown model parameters (μ, σ^2, τ^2).

   Let:

$$\begin{aligned} \mu &\sim \mathrm{N}(\bar{\mathbf{y}}, \sigma_0^2) \\ \sigma^2 &\sim \mathrm{IG}(a_1, b_1) \\ \tau^2 &\sim \mathrm{IG}(a_2, b_2) \end{aligned}$$

   Let n be the total number of observations, n_i be the number of observations in school i, and p be the number of schools, then:

$$\begin{aligned}
\sigma^2|\cdots &\sim \mathrm{IG}\left(\frac{n}{2} + \frac{p}{2} + a_1, \frac{\sum_{j=1}^{n}(\theta_i - y_{ij})^2}{2} + \frac{\sum_{i=1}^{p}(\theta_i - \mu)^2}{2\tau^2} + b_1\right) \\
\tau^2|\cdots &\sim \mathrm{IG}\left(\frac{p}{2} + a_2, \frac{\sum_{i=1}^{p}(\theta_i - \mu)^2}{2\sigma^2} + b_2\right) \\
\mu|\cdots &\sim \mathrm{N}\left(\left(\frac{\sum_{i=1}^{p}\theta_i}{\sigma^2\tau^2} + \frac{\bar{\mathbf{y}}}{\sigma_0^2}\right)\left(\frac{p}{\sigma^2\tau^2} + \frac{1}{\sigma_0^2}\right)^{-1}, \left(\frac{p}{\sigma^2\tau^2} + \frac{1}{\sigma_0^2}\right)^{-1}\right) \\
\theta_i|\cdots &\sim \mathrm{N}\left(\left(\frac{\sum_{j=1}^{n_i} y_{ij}}{\sigma^2} + \frac{\mu}{\sigma^2\tau^2}\right)\left(\frac{n_i}{\sigma^2} + \frac{1}{\sigma^2\tau^2}\right)^{-1}, \left(\frac{n_i}{\sigma^2} + \frac{1}{\sigma^2\tau^2}\right)^{-1}\right)
\end{aligned}$$

For my Gibbs sampler, I chose $\sigma_0^2 = 10$ and $a_1 = b_1 = a_2 = b_2 = 1$. As you can see in the plot below, the hierarchical model shrinks the estimated posterior means for each school towards the total mean.

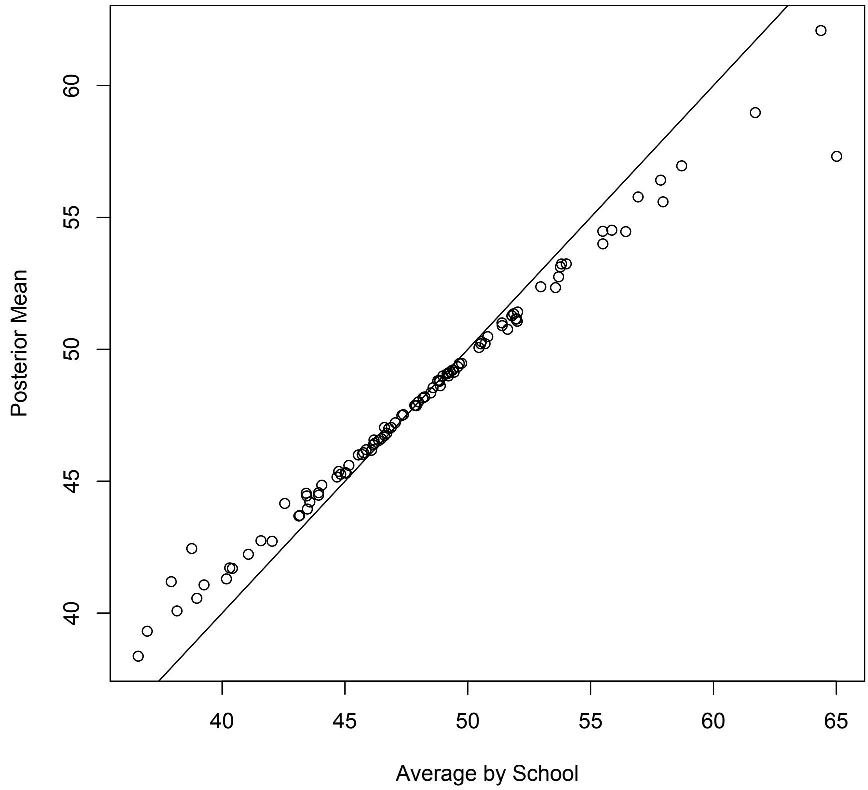

Figure 1: Plot of the School Average by the Estimated Posterior Mean of the School

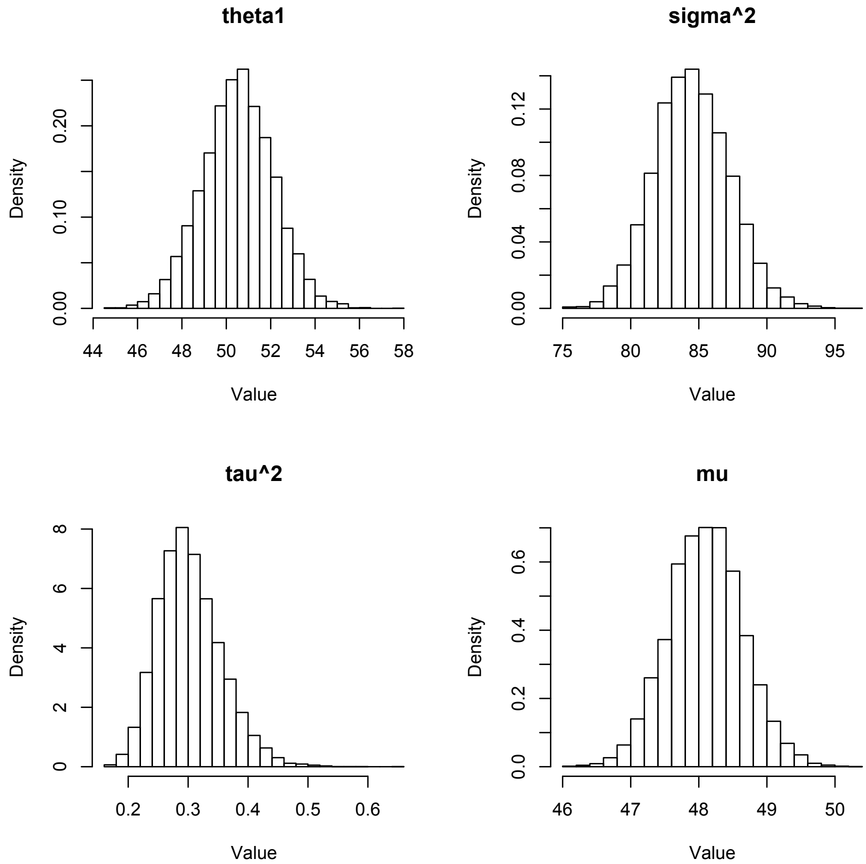

Figure 2: Histograms of Some of the Estimated Coefficients

3. Suppose you use the posterior mean $\hat{\theta}_i$ from the above model to estimate each school-level mean θ_i. Define the shrinkage coefficient κ_i as

$$\kappa_i = \frac{\bar{y}_i - \hat{\theta}_i}{\bar{y}_i} \,,$$

which tells you how much the posterior mean shrinks the observed sample mean. Plot this shrinkage coefficient (in absolute value) for each school as a function of that school's sample size, and comment.

This plot shows us that schools with smaller sample sizes are being shrunk more than schools with larger sample sizes, which is what we would expect.

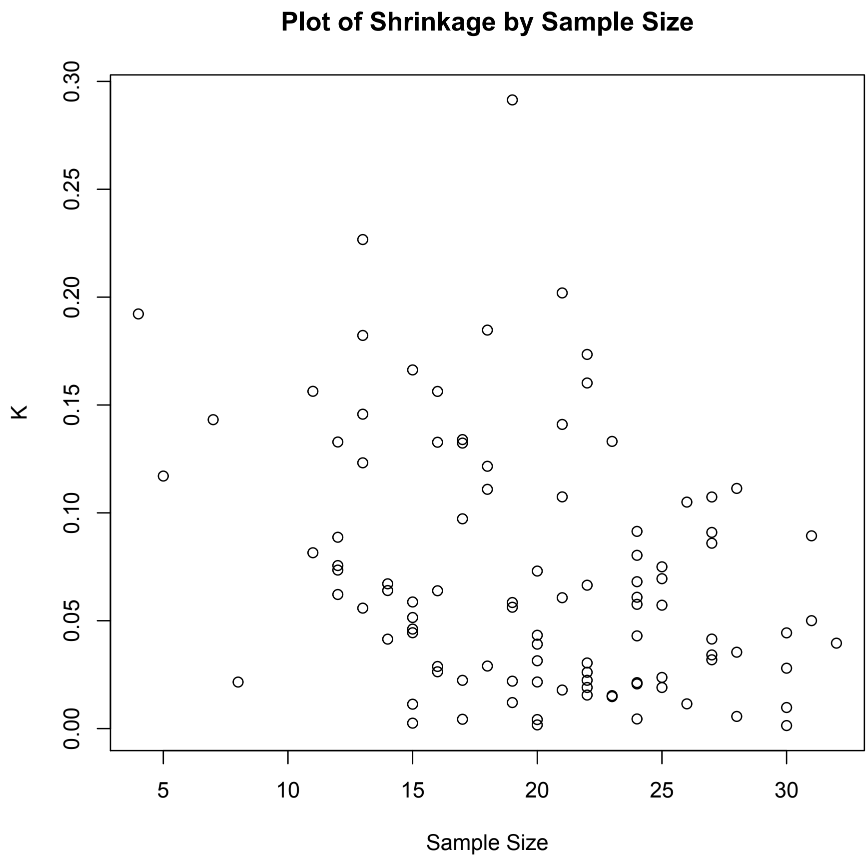

Figure 3: Plot of the shrinkage coefficient κ