# Hierarchical modeling: Normal data

Data and EDA

We have data from the 2002 Educational Longitudinal Study (ELS), a survey of students from a large sample of schools across the United States. This dataset includes a population of schools as well as a population of students within each school. *Note that this is a nested/hierarchical structure.* This survey included 10th grade children from 100 different large urban public high schools, all having a 10th grade enrollment of 400 or greater. Note that the number of students sampled from each school differs.

In particular, we have data from math scores from a math exam on the ELS, which was standardized to produce a nationwide mean of 50 and a standard deviation of 10.

The following displays boxplots of the distribution of math scores by school. The pink line displays the overall sample mean score, averaged across all schools and students. What do we notice?
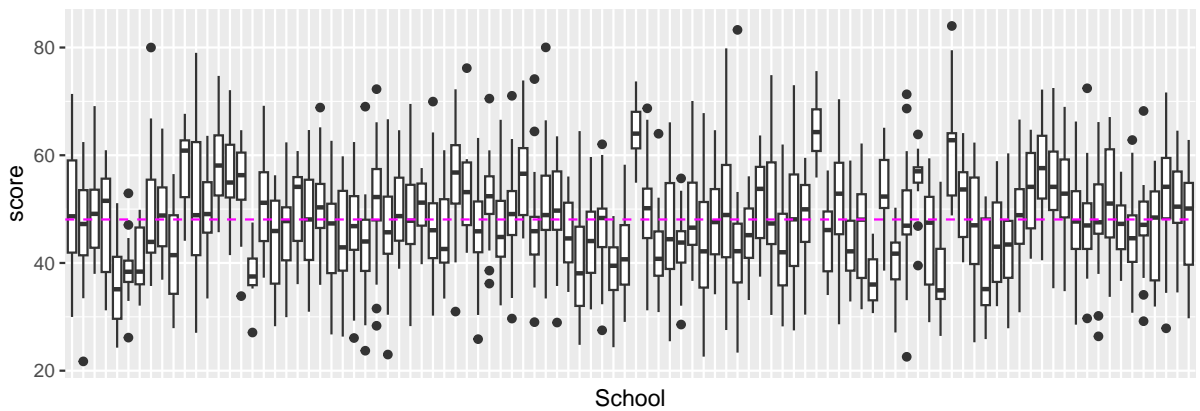
Pink line is overall sample mean

1

Here, we plot the distribution sample mean scores of each school (left), along with the sample mean scores as a function of the sample size from the school (right). What do we notice here?
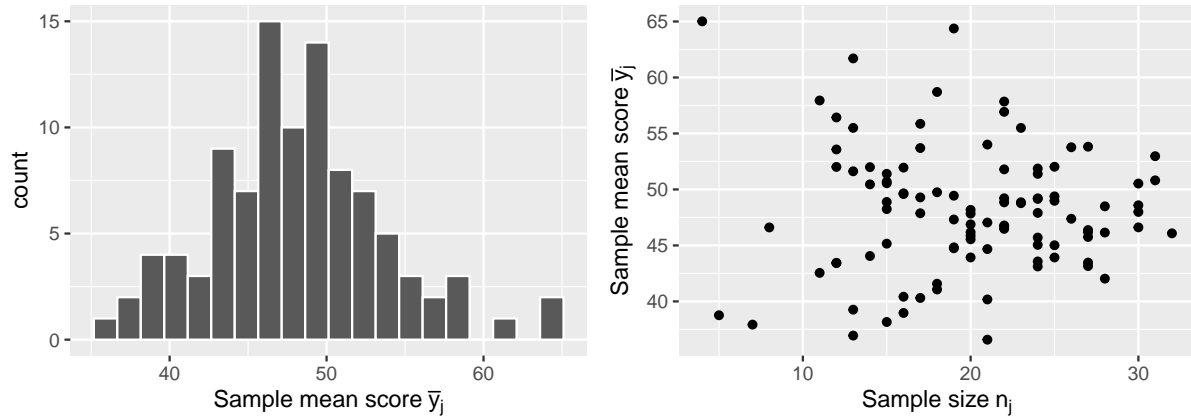

Prior solicitation

Set priors based on the information provided by ELS: the math exam was designed to give a mean of 50 and a nationwide variance of 100. *Note: this variance includes both within-school and between-school variance.*

Let's brainstorm how to translate these into sensible prior parameter values for the unknown parameters! We need to put things in context. Recall:

$$\theta \sim N(\mu_0, \sigma_0^2)$$
$$1/\sigma^2 \sim \text{Gamma}(a, b)$$
$$1/\tau^2 \sim \text{Gamma}(c, d)$$

Some MCMC diagnostics

Ran chain with burn-in of 5000 and then retained 5000 samples. The following plots are another way to determine stationarity/convergence. We can produce boxplots of sequential groups of MCMC samples. In the following, each of the 10 boxplots represents 1/10th of the MCMC samples. If convergence has been achieved, then the distribution of samples in any one boxplot should be the same as that in any other (for a given parameter).
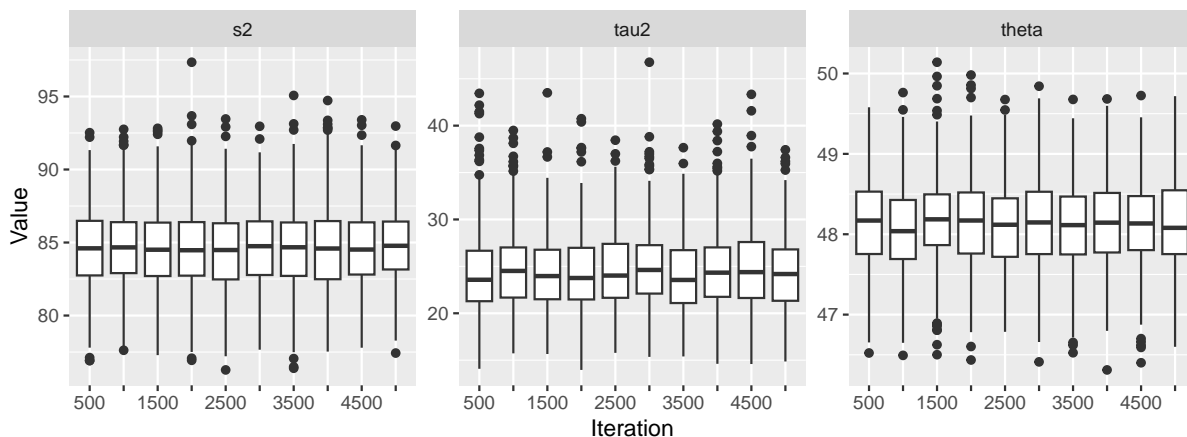
Showing a subset of the θ_j (first 9 schools) for illustrative purposes:
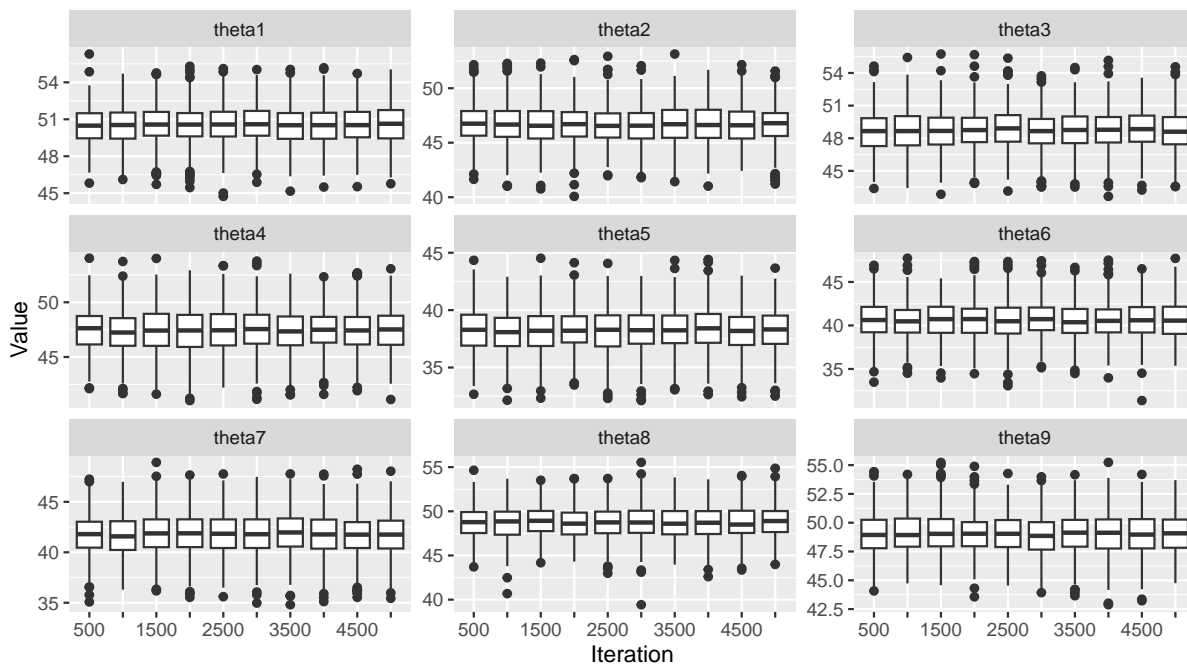

Effective sample sizes from 5000 iterations post burn-in for θ, σ^2, τ^2 are: 3635.62, 4844.69, 2560.76. The ESS for the θ_j chains range from 4121.44 to 5529.28.

Posterior inference

Posterior means and 95% credible intervals of θ, σ, and τ are as follows. Let's think about their interpretation.

	Posterior mean	2.5%	97.5%
θ	48.125	47.024	49.164
σ	9.199	8.912	9.505
τ	4.932	4.186	5.777

The following displays the posterior distributions of the school-specific mean scores θ_j, in order of smallest to largest:

Posterior distribution
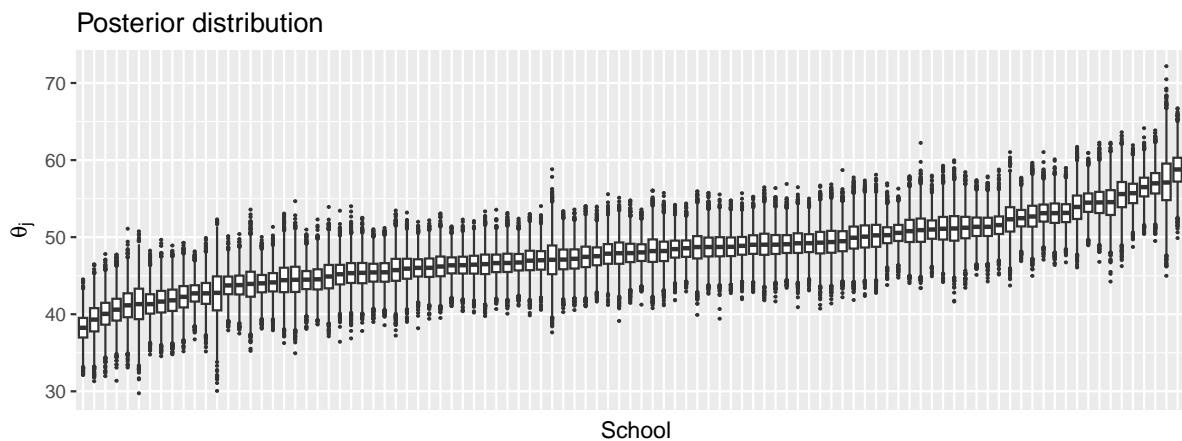
This plot illustrates the notion of "shrinkage" or "borrowing information". Let's interpret!

Shrinkage

4