

**SDS 383D Ex 04:**  
**Hierarchical Models**

March 22, 2016

**Jennifer Starling**

## 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  $\theta_i$  be the underlying mean test score for school  $i$ , and let  $y_{ij}$  be the score for the  $j$ th 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.

### Part 1

Briefly explain why this would be.

The extreme school-level averages occur in the schools with smaller sample sizes because we do not do a very good job of estimating the mean when sample size is small. These schools do not have min and max observation values that are more extreme than the other schools; they just have fewer observations to balance out the calculation of the mean. The smaller the sample size, the more influential an extreme observation is over the group mean.

### Part 2

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

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

Decide upon sensible priors for the unknown model parameters  $(\mu, \sigma^2, \tau^2)$ .

The model is as follows.

$$\begin{aligned} (y_{ij} | \theta_i, \sigma^2) &\sim N(\theta_i, \sigma^2) \\ (\theta_i | \mu, \sigma^2, \tau^2) &\sim N(\mu, \sigma^2 \tau^2) \\ \mu &\sim I_{\mathbb{R}}(\mu), \text{ a flat prior on the real line} \\ \tau^2 &\sim I_{\mathbb{R}^+}(\tau^2), \text{ a flat prior on the positive real line} \\ \sigma^2 &\sim \left( \frac{1}{\sigma^2} \right) I_{\mathbb{R}^+}(\sigma^2), \text{ Jeffreys prior} \end{aligned}$$

where

- $i = 1, \dots, p$  indexes the  $p$  groups.
- $n_i$  = sample size in each group.
- $j = 1, \dots, n_i$  indexes observations in a group.
- $n$  = total number of observations.

The likelihood is

$$L(y | \theta_1, \dots, \theta_p, \sigma^2) \sim \prod_{i=1}^p \prod_{j=1}^{n_i} \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2}} \exp \left[ -\frac{1}{\sigma^2} (y_{ij} - \theta_i)^2 \right] = (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 \right]$$

The full conditionals are as follows.

---


$$(\theta_i | y, \mu, \sigma^2, \tau^2)$$

Note that  $\bar{y}_i$  is a sufficient statistic for the  $y$ 's, with  $\bar{y}_i \sim N\left(\theta_i, \frac{\sigma^2}{n}\right)$ .

$$(\theta_i | y, \mu, \sigma^2, \tau^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2/n}(\bar{y}_i - \theta_i)^2\right] (\tau^2 \sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2 \tau^2}(\theta_i - \mu)^2\right]$$

This is the normal-normal model, therefore

$$(\theta_i | y, \mu, \sigma^2, \tau^2) \sim N(m^*, v^*)$$

with

$$\begin{aligned} v^* &= \left[ \frac{n_i}{\sigma^2} + \frac{1}{\sigma^2 \tau^2} \right]^{-1} = \left[ \frac{n_i \tau^2 + 1}{\sigma^2 \tau^2} \right]^{-1} = \sigma^2 \left[ \frac{\tau^2}{n_i \tau^2 + 1} \right] \\ m^* &= v^* \left[ \left( \frac{n_i}{\sigma^2} \right) \bar{y}_i + \left( \frac{1}{\sigma^2 \tau^2} \right) \mu \right] \\ &= \sigma^2 \left[ \frac{\tau^2}{n_i \tau^2 + 1} \right] \left[ \left( \frac{n_i}{\sigma^2} \right) \bar{y}_i + \left( \frac{1}{\sigma^2 \tau^2} \right) \mu \right] \\ &= \left[ \frac{n_i \tau^2}{n_i \tau^2 + 1} \right] \bar{y}_i + \left[ \frac{1}{n_i \tau^2 + 1} \right] \mu \\ &= w \bar{y}_i + (1 - w) \mu \end{aligned}$$

So full conditional is

$$(\theta_i | y, \mu, \sigma^2, \tau^2) \sim N\left(\left[ \frac{n_i \tau^2}{n_i \tau^2 + 1} \right] \bar{y}_i + \left[ \frac{1}{n_i \tau^2 + 1} \right] \mu, \sigma^2 \left[ \frac{\tau^2}{n_i \tau^2 + 1} \right]\right) \quad (1)$$

---


$$(\mu | \theta, y, \sigma^2, \tau^2)$$

$$\begin{aligned} (\mu | \theta, y, \sigma^2, \tau^2) &\propto \exp\left[-\frac{1}{2\sigma^2 \tau^2} \sum_{i=1}^p (\theta_i - \mu)^2\right] \cdot 1 \\ &= \exp\left[-\frac{1}{2\sigma^2 \tau^2} \{(\theta_1 - \mu)(\theta_1 - \mu) + \dots + (\theta_p - \mu)(\theta_p - \mu)\}\right] \\ &= \exp\left[-\frac{1}{2\sigma^2 \tau^2} \left\{ p\mu^2 - 2\mu \sum_{i=1}^p \theta_i + \sum_{i=1}^p \theta_i^2 \right\}\right] \\ &= \exp\left[-\frac{p}{2\sigma^2 \tau^2} \left\{ \mu^2 - 2\mu \left( \frac{\sum_{i=1}^p \theta_i}{p} \right) + \frac{\sum_{i=1}^p \theta_i^2}{p} \right\}\right] \\ &\propto \exp\left[-\frac{p}{2\sigma^2 \tau^2} \left\{ \mu^2 - 2\mu \bar{\theta}_i \right\}\right] \end{aligned}$$

We recognize this as a Normal kernel, therefore

$$(\mu | \theta, y, \sigma^2, \tau^2) \sim N\left(\bar{\theta}_i, \frac{\sigma^2 \tau^2}{p}\right) \quad (2)$$

---


$$(\sigma^2 | \theta, y, \mu, \tau^2)$$

$$\begin{aligned}
(\sigma^2 | \theta, y, \mu, \tau^2) &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 \right] (\sigma^2)^{-\frac{p}{2}} \exp \left[ -\frac{1}{2\sigma^2 \tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right] \left( \frac{1}{\sigma^2} \right) \\
&= (\sigma^2)^{-\frac{(n+p)}{2}-1} \exp \left[ -\left( \frac{1}{\sigma^2} \right) \cdot \left\{ \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 + \frac{1}{2\tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right\} \right]
\end{aligned}$$

We recognize this as an Inverse-Gamma kernel, therefore

$$(\sigma^2 | \theta, y, \mu, \tau^2) \sim IG \left( \frac{(n+p)}{2}, \left\{ \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 + \frac{1}{2\tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right\} \right) \quad (3)$$

$$(\tau^2 | \theta, y, \mu, \sigma^2)$$

$$(\tau^2 | \theta, y, \mu, \sigma^2) \propto (\tau^2)^{-\frac{p}{2}} \exp \left[ -\frac{1}{2\sigma^2 \tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right] \cdot 1$$

We recognize this as an Inverse Gamma kernel, therefore

$$(\tau^2 | \theta, y, \mu, \sigma^2) \sim IG \left( \frac{p}{2} - 1, \frac{1}{2\sigma^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right) \quad (4)$$

### Part 3

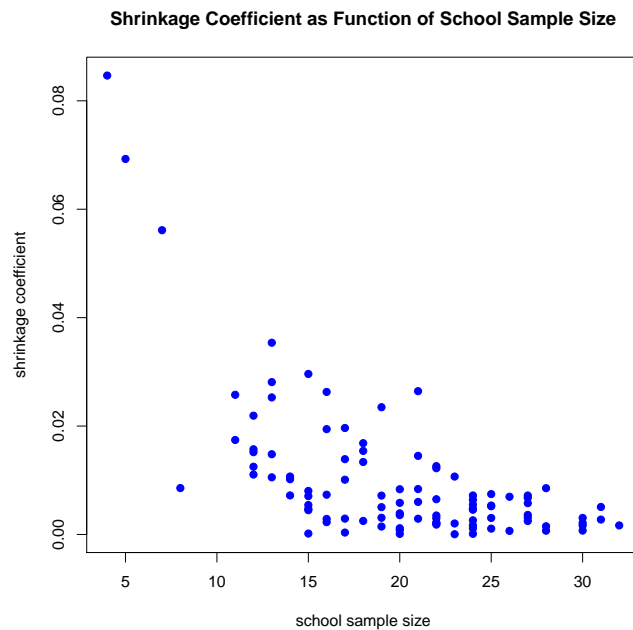


Figure 1: Shrinkage estimator by school sample size

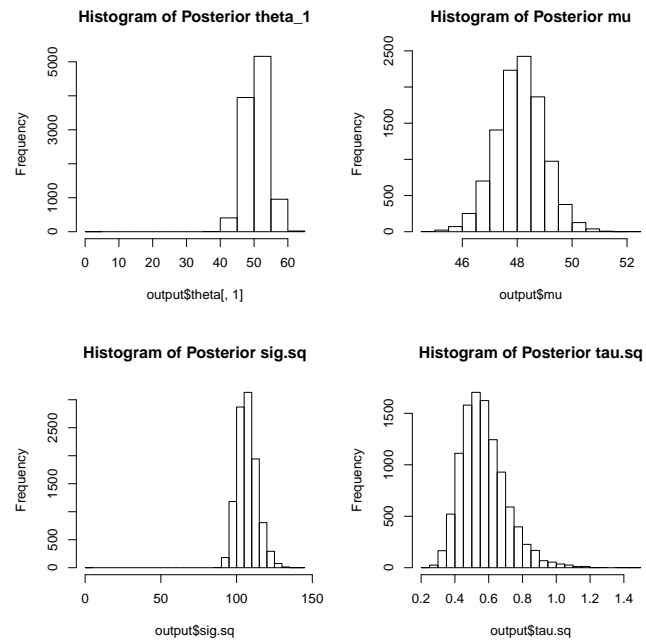


Figure 2: Histograms of Posteriors

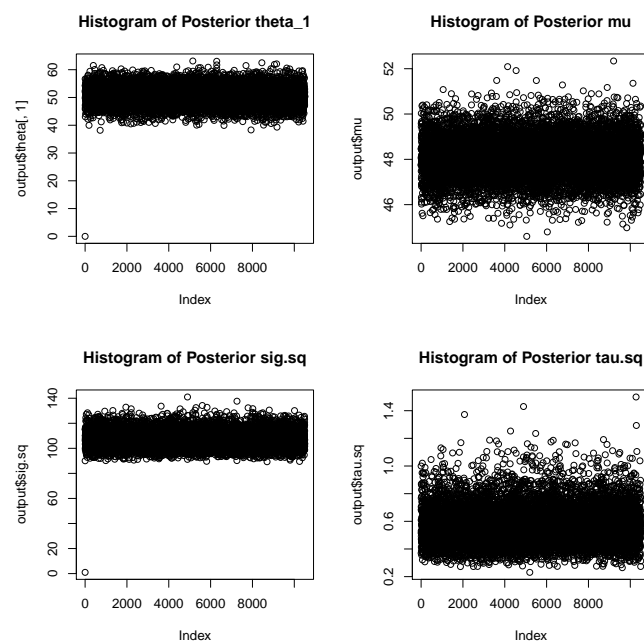


Figure 3: Traces for Gibbs Sampler

## Price Elasticity of Demand

### Linear Hierarchical Model using Empirical Bayes

Model is specified as

$$\begin{aligned}
 \log(Q_{it}) &= \log(\alpha_i) + \beta_i \log(P_{it}) + \gamma_i x_{it} + \theta_i (\log(P_{it}) * x_{it}) + e_{it} \\
 \alpha_i &\sim N(\mu_\alpha, \tau_\alpha^2) \\
 \beta_i &\sim N(\mu_\beta, \tau_\beta^2) \\
 \gamma_i &\sim N(\mu_\gamma, \tau_\gamma^2) \\
 \theta_i &\sim N(\mu_\theta, \tau_\theta^2) \\
 e_{it} &\sim N(0, \sigma^2)
 \end{aligned}$$

Where  $i = \{1, 2, \dots, 88\}$  indexes stores, and  $t = \{1, 2, \dots, 68\}$  indexes week (repeated obs on each store).

$\log(Q_{it})$  = Response; log-volume for store  $i$  at week  $t$

$\log(P_{it})$  = Log-price for store  $i$  at week  $t$

$\log(\alpha_i)$  = Intercept for each store

$x_{it}$  = Indicator variable for ad display (displayed ad = 1)

$\log(P_{it}) * x_{it}$  = Interaction; shape may change depending on whether ad in store

Variance estimates using lmer to fit the model were as follows.

$$\hat{\tau}_\alpha^2 = 1.21606$$

$$\hat{\tau}_\beta^2 = 0.54451$$

$$\hat{\tau}_\gamma^2 = 0.93347$$

$$\hat{\tau}_\theta^2 = 0.78085$$

$$\hat{\sigma}^2 = 0.06733$$

Residual plot does not show evidence of major model mis-fit.

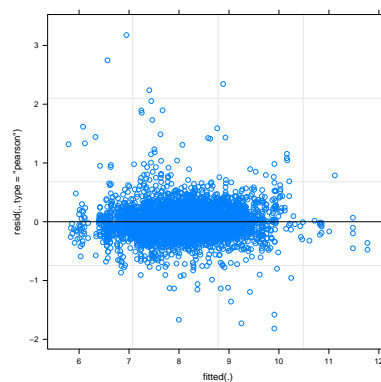


Figure 4: Residual plot for hierarchical model

Model summary:

```

Linear mixed model fit by REML ['lmerMod']
Formula: logQ ~ logP + disp + disp:logP + (1 + logP + disp + disp * logP | store)
Data: data

5 REML criterion at convergence: 1664.9

Scaled residuals:
    Min      1Q  Median      3Q      Max
-7.0011 -0.4879 -0.0247  0.4346 12.2396

10 Random effects:
    Groups      Name      Variance Std.Dev. Corr
store (Intercept) 1.21606   1.1028
      logP        0.54451   0.7379  -0.74
15      disp        0.93347   0.9662  -0.12  0.04
      logP:disp    0.78085   0.8837   0.15 -0.11 -0.99
Residual        0.06733   0.2595
Number of obs: 5555, groups: store, 88

20 Fixed effects:
      Estimate Std. Error t value
(Intercept) 10.2399    0.1479  69.24
logP         -2.1228    0.1166 -18.20
disp          0.4387    0.1433   3.06
25 logP:disp   -0.2355    0.1336  -1.76

Correlation of Fixed Effects:
      (Intr) logP  disp
logP   -0.840
30 disp  -0.415  0.431
logP:disp 0.428 -0.472 -0.989

```

**Fully Bayesian Hierarchical Linear Model**

## **A Hierarchical Probit Model via Data Augmentation**



## Gene Expression Over Time