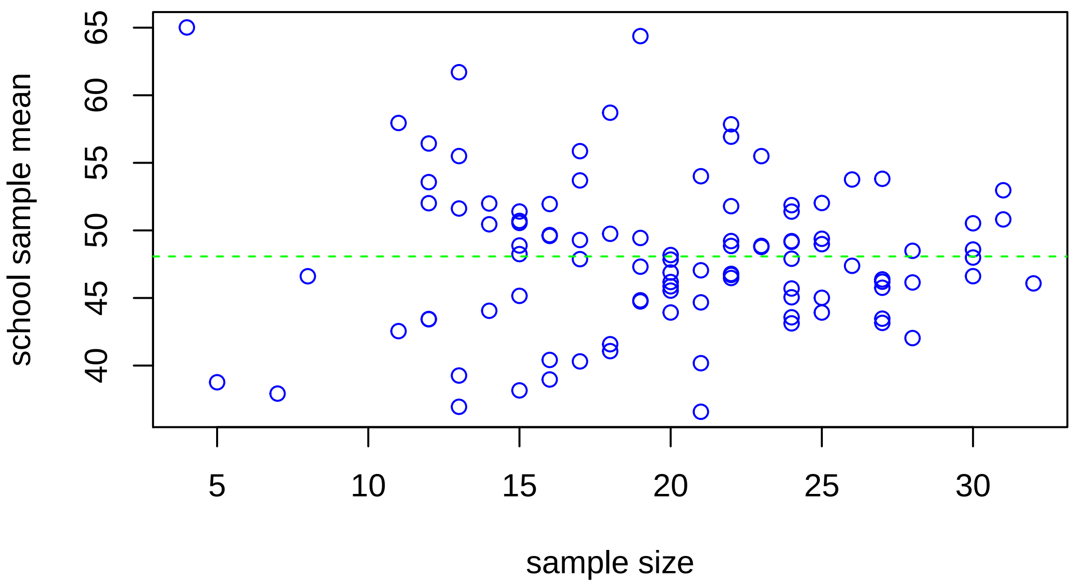
Math Tests

1. With smaller sample sizes, the standard deviation of the sampling distribution is larger, so we expect the sample means with smaller sample sizes to be farther from their respective means than those associated with larger sample sizes. We should not be surprised, therefore, that the smaller sample sizes also tend to lead to sample means farther from the grand mean, which is shown by the green dashed line. The following plot shows this:



2. Let $i \in \{1 \ldots n\}$ denote the school, $j \in \{1, \ldots, m_i\}$ the observations from school i, and M the total number of observations. The model is as follows:

$$
\begin{aligned}
y_{ij} &\sim N\left(\theta_i, \sigma^2\right) \\
\theta_i &\sim N\left(\mu, \tau^2\sigma^2\right) \\
&\left.\begin{aligned}
\mu &\sim 1 \\
\sigma^2 &\sim 1/\sigma^2 \\
\tau^2 &\sim 1
\end{aligned}\right\} \text{improper}
\end{aligned}
$$

The prior for σ^2 is Jeffrey's prior, as we saw in class. The prior for μ is flat, which we also saw in class. I chose a flat prior for τ^2, as well, after skimming the Gelman article. I decided to use that instead of an inverse gamma since, as we shall see, the flat prior updates to an inverse gamma anyway, and I chose a simpler non-informative prior to avoid choosing hyper-hyperparameters. With this setup, the joint posterior distribution, with $\theta := \langle \theta_1, \ldots, \theta_n \rangle$ is

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

We can now write the conditional distributions. As we saw in class,

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

We can get the others with the usual methods.

$$
f(\sigma^2 | \ldots) \propto \left(\frac{1}{\sigma^2} \right)^{(M+n)/2+1} \exp\left(-\frac{1}{2\sigma} \left(\sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \theta_i)^2 + \sum_{i=1}^{n} \frac{(\theta_i - \mu)^2}{\tau^2} \right) \right)
$$

whence

$$
\sigma^2 | \ldots \sim \mathrm{IG}\left(\frac{M+n}{2}, \quad \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \theta_i)^2 + \sum_{i=1}^{n} \frac{(\theta_i - \mu)^2}{\tau^2} \right) \right)
$$

To get the conditional distribution for μ, let us first note that

$$
\begin{aligned}
\sum_{i=1}^{n} (\theta_i - \mu)^2 &= \sum_{i=1}^{n} (\mu^2 - 2\mu\theta_i + \theta_i^2) \\
&= n\mu^2 - 2n\mu\bar{\theta} + K
\end{aligned}
$$

for some K not involving μ. Thus,

$$
\begin{aligned}
f(\mu | \ldots) &\propto \exp\left(-\frac{1}{2\sigma^2\tau^2} \sum_{i=1}^{n} (\mu - \theta_i)^2 \right) \\
&\propto \exp\left(-\frac{n}{\sigma^2\tau^2} (\mu^2 - 2\mu\bar{\theta}) \right)
\end{aligned}
$$

whence

$$
\mu | \ldots \sim N\left(\bar{\theta}, \frac{\sigma^2\tau^2}{n} \right)
$$

Finally,

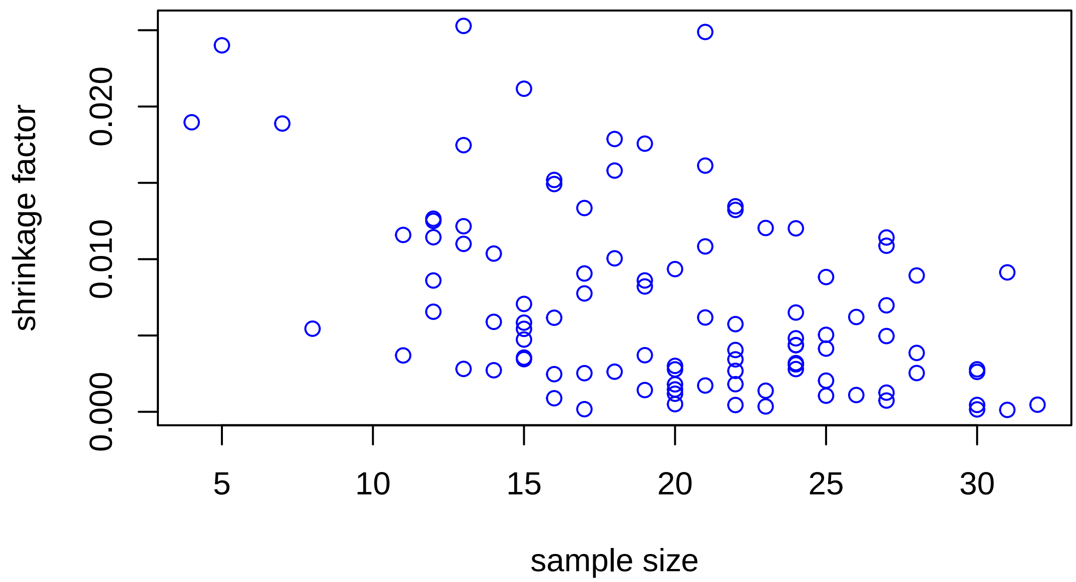
$$
f(\tau^2 | \ldots) \propto \left(\frac{1}{\tau^2} \right)^{n/2} \exp\left(-\frac{1}{2\sigma^2\tau^2} \sum_{i=1}^{n} (\theta_i - \mu)^2 \right)
$$

whence

$$
\tau^2 | \ldots \sim \mathrm{IG}\left(\frac{n}{2}, \sum_{i=1}^{n} \frac{(\theta_i - \mu)^2}{2\sigma^2} \right)
$$

With these distributions, we can set up a Gibbs sampler to sample from the posterior distribution. (See the file named Math.r.)

3. After running the Gibbs sampler, we get a sample for the posterior estimation for each θ_i. From this we can define the shrinkage factors $(y_{ij} - \hat{\theta}_i)/y_{ij}$, whose absolute values are plotted below.



Price elasticity of demand

Let $i \in \{1, \ldots, m := 88\}$ index the stores, and let n_i denote the number of observations for store i. Let n denote the total number of observations; that is, $n := \sum_{i=1}^{m} n_i$. For each i, let X_i be the $n_i \times 4$ matrix with an intercept, the log-price, display marker, and price/display interaction term for each observation of store i. We shall create a model with unmodeled and modeled effects as follows.

$$\begin{aligned}
y_i &\sim N\big(X_i\beta + X_i\gamma_i, \sigma^2 I\big), \text{ where } y_i \text{ is the log-volume vector for store } i \\
\sigma^2 &\sim 1/\sigma^2, \text{ where we use } \lambda := 1/\sigma^2 \text{ when convenient} \\
\beta &\sim N(\mu, \Sigma), \text{ where } \beta \text{ is the coefficient vector for overall effects (not store specific)} \\
\gamma_i &\sim N(0, T), \text{ where } \gamma_i \text{ is the coefficient vector for store } i\text{'s offset from the average effects} \\
T &\sim IW(\nu, V)
\end{aligned}$$

The hyperparameters for β and T are unmodeled. To run a Gibbs sampler, we need the full conditionals for the modeled (hyper)parameters.

1. For λ, we need the full likelihood (using all stores's data), which is proportional to

$$\lambda^{n/2} \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{m}\Big(y_i - X_i\beta - X_i\gamma_i\Big)^{\mathsf{T}}\Big(y_i - X_i\beta - X_i\gamma_i\Big)\right)$$

Multiplying this by the prior gives us

$$f(\lambda|\ldots) \propto \lambda^{n/2-1} \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{m}\Big(y_i - X_i\beta - X_i\gamma_i\Big)^{\mathsf{T}}\Big(y_i - X_i\beta - X_i\gamma_i\Big)\right)$$

which is a gamma density, and thus we see that

$$\sigma^2|\ldots \sim IG\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{m}\Big(y_i - X_i\beta - X_i\gamma_i\Big)^{\mathsf{T}}\Big(y_i - X_i\beta - X_i\gamma_i\Big)\right)$$

2. For each γ_i, we need the store-specific likelihood function, which is proportional to

$$\exp\left(-\frac{\lambda}{2}\Big(y_i - X_i\beta - X_i\gamma_i\Big)^{\mathsf{T}}\Big(y_i - X_i\beta - X_i\gamma_i\Big)\right)$$

Thus, the conditional posterior density for γ_i is

$$\begin{aligned}
f(\gamma_i|\ldots) &\propto \exp\left(-\frac{1}{2}\gamma_i^{\mathsf{T}}T^{-1}\gamma_i\right)\exp\left(-\frac{\lambda}{2}\Big(y_i - X_i\beta - X_i\gamma_i\Big)^{\mathsf{T}}\Big(y_i - X_i\beta - X_i\gamma_i\Big)\right) \\
&\propto \exp\left(-\frac{1}{2}\gamma_i^{\mathsf{T}}T^{-1}\gamma_i - \frac{\lambda}{2}\Big(\gamma_i^{\mathsf{T}}X_i^{\mathsf{T}}X_i\gamma_i - 2\gamma_i^{\mathsf{T}}\big(X_i^{\mathsf{T}}y_i - X_i^{\mathsf{T}}X_i\beta\big)\Big)\right) \\
&= \exp\left(\gamma_i^{\mathsf{T}}\Big(T^{-1} + \lambda X_i^{\mathsf{T}}X_i\Big)\gamma_i - 2\gamma_i^{\mathsf{T}}\lambda\Big(X_i^{\mathsf{T}}y_i - X_i^{\mathsf{T}}X_i\beta\Big)\right).
\end{aligned}$$

This is proportional to a normal density; in particular,

$$\gamma_i|\ldots \sim N(\kappa^*, T^*),$$

where

$$\begin{aligned}
T^* &:= \big(T^{-1} + \lambda X_i^{\mathsf{T}}X_i\big)^{-1} \\
\kappa^* &:= \lambda T^*\big(X_i^{\mathsf{T}}y_i - X_i^{\mathsf{T}}X_i\beta\big)
\end{aligned}$$

3. To get the conditional for β, we follow a pattern similar to the simplification seen above for γ_i, but with the roles of β and γ_i switched and a summation over all stores involved. We therefore see that the likelihood is, up to proportionality, equal to

$$\exp\left(-\frac{\lambda}{2}\sum_{i=1}^{m}\left(\beta^\intercal X_i^\intercal X_i\beta - 2\beta^\intercal\left(X_i^\intercal y_i - X_i^\intercal X_i\gamma_i\right)\right)\right)$$

The prior for β reduces by proportionality to

$$\exp\left(-\frac{1}{2}\left(\beta^\intercal\Sigma^{-1}\beta - 2\beta^\intercal\Sigma^{-1}\mu\right)\right)$$

We thus get

$$\begin{aligned}
f(\beta|\ldots) &\propto \exp\left(-\frac{1}{2}\left(\beta^\intercal\Sigma^{-1}\beta - 2\beta^\intercal\Sigma^{-1}\mu\right)\right)\exp\left(-\frac{\lambda}{2}\sum_{i=1}^{m}\left(\beta^\intercal X_i^\intercal X_i\beta - 2\beta^\intercal\left(X_i^\intercal y_i - X_i^\intercal X_i\gamma_i\right)\right)\right)\\
&= \exp\left[-\frac{1}{2}\left(\beta^\intercal\left(\Sigma^{-1} + \lambda\sum_{i=1}^{m}X_i^\intercal X_i\right)\beta - 2\beta^\intercal\left(\Sigma^{-1}\mu + \lambda\sum_{i=1}^{m}\left(X_i^\intercal y_i - X_i^\intercal X_i\gamma_i\right)\right)\right)\right]
\end{aligned}$$

This is the kernel of a multivariate normal density, so

$$\beta|\ldots \sim N(\mu^*, \Sigma^*),$$

where

$$\begin{aligned}
\Sigma^* &:= \left(\Sigma^{-1} + \lambda\sum_{i=1}^{m}X_i^\intercal X_i\right)^{-1}\\
\mu^* &:= \Sigma^*\left(\Sigma^{-1}\mu + \lambda\sum_{i=1}^{m}\left(X_i^\intercal y_i - X_i^\intercal X_i\gamma_i\right)\right)
\end{aligned}$$

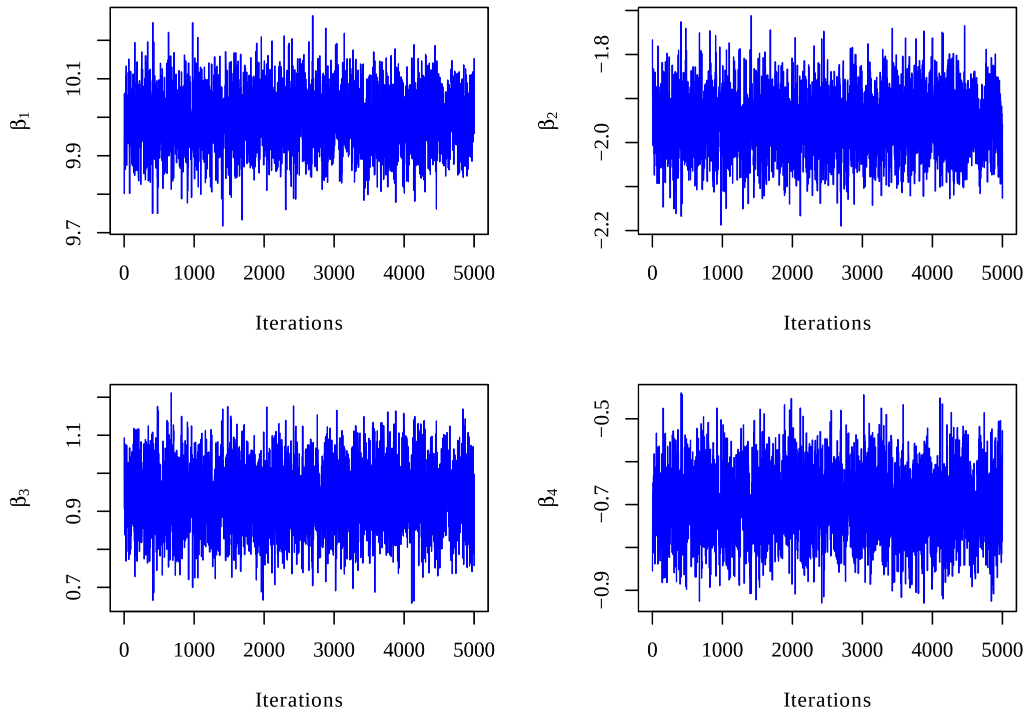
4. For T, first recall the identity $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ whenever A and B are conformable for both products. Recall also that $x^\intercal Ax = \mathrm{tr}(x^\intercal Ax)$ for any quadratic form $x^\intercal Ax$. We need only the part of the full likelihood that involves T, and we combine that with the prior to get

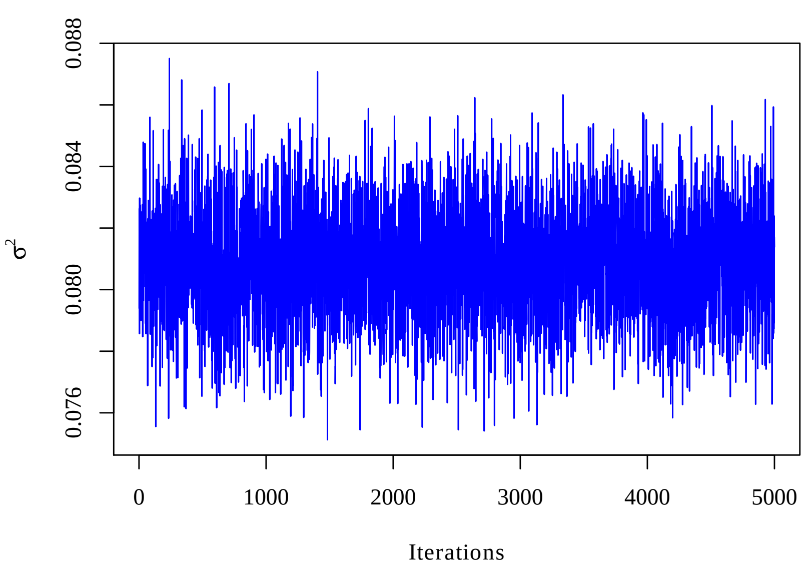
$$\begin{aligned}
f(T|\ldots) &\propto |T|^{-(\nu+3)/2}\exp\left(-\frac{1}{2}\mathrm{tr}\left(VT^{-1}\right)\right)|T|^{-m/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\gamma_i^\intercal T^{-1}\gamma_i\right)\\
&= |T|^{-(\nu+m+3)/2}\exp\left(-\frac{1}{2}\mathrm{tr}\left(VT^{-1}\right) - \frac{1}{2}\sum_{i=1}^{m}\gamma_i^\intercal T^{-1}\gamma_i\right)\\
&= |T|^{-(\nu+m+3)/2}\exp\left(-\frac{1}{2}\mathrm{tr}\left(VT^{-1}\right) - \frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{m}\gamma_i\gamma_i^\intercal T^{-1}\right)\right)\\
&= |T|^{-(\nu+m+3)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\left(V + \sum_{i=1}^{m}\gamma_i\gamma_i^\intercal\right)T^{-1}\right)\right]
\end{aligned}$$

and therefore

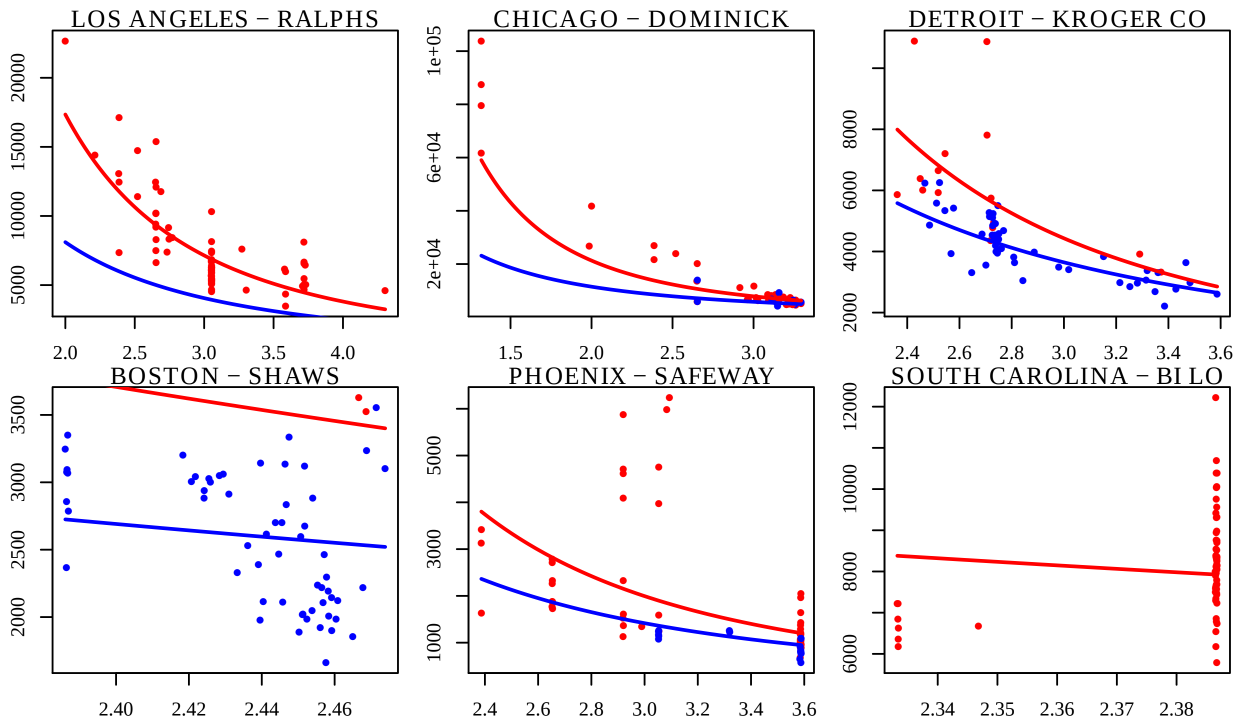
$$T|\ldots \sim IW\left(\nu + m,\ V + \sum_{i=1}^{m}\gamma_i\gamma_i^\intercal\right).$$

After running the Gibbs sampler, we can graph the samples for the parameters. Shown below are the plots for β and σ^2.





Putting the pieces together, we can graph the modeled demand curves. Below are six such curves.



There is shrinkage toward a common mean, but this is not completely evident here.

To see this, let us instead plot the log demand curves for four stores. Shown first are the curves as estimated by ordinary least squares, and second those estimated by our model.