I collected the counts of the number of visitors to my Baseball/R blog for 20 days in May and June, 2016.

84, 75, 106, 138, 87, 44, 47, 122, 201, 118, 93, 84, 63, 59, 72, 114, 111, 72, 69, 59

A reasonable sampling model is the negative binomial with density

\[ f(y|\mu, n) = \frac{\Gamma(y + n)}{\Gamma(n) y!} p^n (1 - p)^y, \]

where \( p = \frac{n}{n + \mu} \).

Assuming a uniform prior on \( \theta = (\log \mu, \log n) \), the posterior density of \( \theta \) is given by

\[ g(\theta|y) \propto \prod_{j=1}^{20} f(y_j|\mu, n) \]
By using some MCMC algorithm, take a simulated sample of size 100 from the posterior distribution of $\theta$.
From the sample, find 90 percent interval estimates for $\mu$ and $n$.
Predict the number of total visitors in 5 future days.
Fit a traditional regression model with normal errors using the \texttt{lm} function. Compare the parameter estimates of $\beta$ with the estimates using the t-error model.

There are alternative heavy-error models that can downweigh unusual observations. One alternative is the logistic error model. Use LearnBayes with a slight change to the log posterior function to fit a regression model with logistic errors. (In R the function \texttt{dlogis} function will compute the logistic pdf.)
There are some unusually large and small worship counts due to special holidays, weather, and other causes. Create a new data frame with these unusual counts removed (in my work, I identified the weeks 15 65 105 114 151 167 218 265 266 267 319 322 358 373 386 411 428 438 463 477 as unusual). Refit the negative-binomial model with new data and contrast the fit with the fit to the original dataset.
Waiting Times at a Coffee Shop

Suppose one focuses on the morning waiting times of the $J$ coffee shops. One considers the “varying intercepts" model

$$y_i \sim N(\mu_j[i], \sigma^2), \quad i = 1, \ldots, N$$

where the intercepts $\mu_1, \ldots, \mu_J$ follow the multilevel model

$$\begin{align*}
\mu_1, \ldots, \mu_J &\sim N(\theta, \tau^2) \\
(\theta, \tau^2) &\sim g(\theta, \tau^2) = 1
\end{align*}$$

(We assume the sample standard deviation $\sigma$ is known.)

1. First simulate data from this model. Assume that $\theta = 5$, $\tau = 1$, there are $J = 20$ coffee shops, and you will have a sample of $n = 5$ waiting times for each shop (so $N = 100$). Assume that the sampling standard deviation is $\sigma = .75$.

2. Explore the following computation strategies to estimate the second-stage parameters $\theta$ and $\tau^2$. 
Strategy One (LearnBayes)

Let $\bar{y}_j$ denote the sample mean of the $j$th group. One can show that the marginal posterior distribution of $(\theta, \log \tau^2)$ is given by

$$g(\theta, \log \tau^2) \propto \prod_{j=1}^{J} \phi \left( \bar{y}_j, \theta, \frac{\sigma^2}{n} + \tau^2 \right) \tau^2$$

Here’s a function to compute the log posterior:

```r
logpost <- function(theta_vector, data){
  theta <- theta_vector[1]
  tausq <- exp(theta_vector[2])
  ybar <- data[, 1]
  sigmasq <- data[, 2]
  sum(dnorm(ybar, theta, sqrt(sigmasq + tausq), 
            log=TRUE)) + log(tausq)
}
```
Strategy One (LearnBayes)

- Use the function laplace in the LearnBayes package to find the posterior mean and standard deviation of $\theta$ and $\log \tau^2$.
- Take a sample of size 1000 from the posterior distribution of $(\theta, \log \tau^2)$.
- Use the simulated sample to find 90 percent interval estimates for $\theta$ and $\tau$. 
Strategy Two (JAGS)

The following JAGS script defines the varying intercepts model. The variable prec.y is the reciprocal of the sampling variance of $\bar{y}_j$ and prec.mu is the reciprocal of $\tau^2$.

```r
modelString = "
model {
  for (i in 1:J){
    y[i] ~ dnorm (mu[i], prec.y)
    mu[i] ~ dnorm(theta, prec.mu)
  }
  prec.mu <- pow(tau2, -1)
  tau2 ~ dunif(0, 100)
  theta ~ dunif(0, 100)
}
writeLines(modelString, con="normexch.bug")
"`
```
Strategy Two (JAGS)

- Use JAGS and this model script to simulate 5000 values from the posterior distribution, collecting values of $\theta$ and $\tau^2$.
- Construct trace plots of the simulated draws of $\theta$ and $\tau^2$ to check convergence of the MCMC chain.
- Use the simulated draws to find 90 percent interval estimates for $\theta$ and $\tau$.
- Compare your results with the results from Strategy One.