Week 2 Homework

1. Using the results in Chapter 2, explore Bayes regression:
   a. simulate from a regression model with k=3. n=30. To calibrate your simulations, pool the cheese data and regress log(VOLUME) on DISP and log(PRICE) to set beta and sigma-squared. Sample from the rows of X for this pooled regression. see the R command `sample` to do this.
   b. Write R code to sample from the prior. Note you will need to sample from a multivariate normal and a Chi-squared. Chi-squared is available as `rchisq` in R; for multivariate normal see Chapter 2 notes.
   c. Plot the prior distribution of beta elements and sigma for various prior settings. Consider the "standard diffuse" prior

$$\sigma^2 \sim \upsilon_0 s_0^2 \Big/ \chi^2_{\upsilon_0} \quad \upsilon_0 = 3;\ s_0^2 = 1$$

$$\beta \Big| \sigma^2 \sim N\left(0, \sigma^2 A^{-1}\right)\ A = .01 I_{\dim(\beta)}$$

      as well as a much tighter prior (change A by a factor of 100!)
   d. What is a possible objection to the prior on σ^2? Hint: look at your simulated data for the standard diffuse prior settings.
   e. Compare the prior and posterior distributions of the model parameters. (here you might find the `boxplot` command useful. If you have data in a matrix X, `boxplot(data.frame(X))` will give you a set of boxplots on the same plotting frame of the columns of X. Use `runireg` to draw from the posterior.
2. The SUR model is discussed in section 2.8.6. Given Σ, there is a normal conjugate prior for this model. Write down the posterior using this prior.
3. In the homework directory, there is a file `rmultireg_wrong.R`. This is R code to simulate from the posterior of the multivariate regression model. There is a conceptual error in this code. Locate the error and explain why it might be hard to detect by pure simulation methods alone. Hints: 1). the error results the wrong value of S being computed. 2). Read section 2.12 carefully.
4. Gelman et al p. 581 suggest that you can simulate from a Wishart distribution with df parameter υ and location S, $W(\upsilon, S)$, by simulating $x_i \sim N(0, S);\ i = 1, ..., \upsilon$ and computing $W = \sum_{i=1}^{\upsilon} x_i x_i'$ Code this up in R. `rwishart` in *`bayesm`* does this by a much more efficient method. Use `rwishart` to check your code for accuracy. Why is the method in `rwishart` and discussed on p. 44 more efficient?