22S:138, Bayesian Statistics
Fall 2007 Homework 8

Due: Mon., 11/12 in class

You will fit a linear regression model to the temperature/heart rate data discussed in class. This dataset is available in WinBUGS format under "Datasets" on the course web page as "normtbug.dat." You may copy it from there into a WinBUGS window. We will consider heart rate the response variable and body temperature the independent or predictor variable.

See the lab 5 handout for example code for fitting a regression model. You will need to change the variable names in the code to match the variable names in the data (i.e. "heart" instead of "Y," etc.).

You will use the noninformative priors from the example code.

1. Fit the model with centering of the predictor variable.

   ```
   mu[i] <- alpha + beta * (temp[i]-temp.bar)
   ```

   Run 3 separate 1000-iteration samplers. Use the following initial values. Monitor alpha, beta, and sigma.

   ```
   list(alpha = -100, beta = 5, tau = 1)
   list(alpha = 0, beta = 2.5, tau = .1)
   list(alpha = 100, beta = 0, tau = 10)
   ```

   Look at history plots of all 3 chains together, as well as autocorrelation plots. Also run the Gelman and Rubin diagnostic. (The only plot you need to print is the history plot for α.)

   Write a paragraph describing the plots and commenting on whether the samplers are converging satisfactorily.

   - How many initial iterations should you throw out in order that the remaining iterations appear to be draws from the target distribution?

2. Print the summary statistics for the model parameters. Have WinBUGS begin computing them starting at the first iteration that you do not throw out. Use the posterior means of α and β to hand-calculate the expected value of *heart* when $temp = 100.8$. Recall that the covariate has been centered.

3. Run the plots in "Inference/Compare." Comment on whether you see outliers, nonlinearity, or unequal variance, etc.

4. Prediction: In the data list, change the last value of "heart" to NA. Then fit the model again. Run the same number of iterations as you threw out in a previous step. Then start monitoring heart[130] as well as alpha, beta, and sigma. Obtain a 95% prediction interval for heart[130].

5. Informative priors: Suppose you had the following SAS output from a linear regression fit to a dataset of body temperatures and heart rates measured on a different random sample of 100 healthy adults. The predictor variable was centered in this analysis.

Model: MODEL1
Dependent Variable: heart

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	651.29310	651.29310	43.93	<.0001
Error	98	1452.80040	14.82449		
Corrected Total	99	2104.09349			

Root MSE	3.85026	R-Square	0.3095
Dependent Mean	77.26720	Adj R-Sq	0.3025
Coeff Var	4.98304		

The REG Procedure
Model: MODEL1
Dependent Variable: heart

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	75.73899	0.60151	122.59	<.0001
temp	1	2.73008	0.41189	6.63	<.0001

(a) Write one choice of informative prior based on this previous analysis.

(b) Would the posterior variance of the regression slope parameter be larger or smaller if you fit a model using this prior rather than the vague prior previously used? Why? (You do not have to refit the model to answer this question.)