HW4 STT 465
(MSU, Fall, 2015)

Due: Friday Nov 20th, 2015 (hard copy on mailbox in Department of Statistics and Probability 619 Red Cedar Road Rm C413)

Note: Please provide a report as neat as you can, with clear answers.

Data. For this HW we will use a mice data set from the Wellcome Trust. The data set provide information on Body Mass Index (BMI, this will be our response) and three factors: litter size, sex and an ID for the cage where the mice were raised. The data set is included with the BGLR package and can be loaded in the R-environment using the following code

```
library(BGLR)
data(mice)
y<-scale(mice.pheno$Obesity.BMI)
sex<-factor(mice.pheno$GENDER)
litterSize<-factor(mice.pheno$Litter)
cage<-factor(mice.pheno$cage)
```

1. Descriptive statistics

Report descriptive statistics for this data, including: (i) a histogram for BMI, (ii) bar-plots of number of records per level of each factor (hint: can use `plot(table(x))`, where x is the variable of interest), and (iii) boxplots of BMI versus each of the factors of interest.

2. Ordinary least squares

Estimate the effects of sex, litter size and cage using OLS. Summarize your findings regarding: (i) whether there are significant effects of sex, and litter size, (ii) whether cage is significant at all, and (iii) model R-squared with and without cage.

Hint: to assess the significance of cage, fit a model with sex and litter size only (null hypothesis) and one with sex, litter size and cage (alternative hypothesis). Then call anova(fm0, fm1) where fm0 and fm1 are the fitted null and alternative hypothesis. The p-value reported correspond to an F-test.

3. Bayesian Analysis

Regress BMI on sex, litter Size and cage using a Bayesian model with normal likelihood and normal prior. In your Bayesian model treat the effects of sex and litter size as fixed and those of cage as random.

Hint: to obtain the design matrices for the different effects you can use the following code

```
XF<-as.matrix(model.matrix(~sex+litterSize))[,-1]
XR<-as.matrix(model.matrix(~cage)[,-1]
XR<-scale(XR,scale=F,center=T)
XF<-scale(XF,scale=F,center=T)
XF<-cbind(1,XF)
```

Note: above we have centered the incidence matrices of effects, except the 1st column of XF which pertains to the intercept.

Report:

- Trace plots for the error variance and for the variance of cage.
- Estimated posterior means and 95% posterior credibility regions for the effects of sex and litter size
- Estimated posterior means and 95% posterior credibility regions for the error variance and the variance of cage effects.

4. Bayesian Vs. OLS

Provide a plot of OLS estimates of cage effects versus Bayesian estimates (estimated posterior means) of the same unknowns. Be sure to have the same range in the x and y axis and add a 45 degree line.

Summarize your findings.

5. Gibbs sampler

Provide in an appendix:

- The function you use to implement the Gibbs sampler
- The code you sue to carry out your analyses.