University of California, Davis
Department of Statistics
Final Project– Spring 2022

Course Title: Bayesian Statistical Inference
Teacher's name: Jairo Fúquene-Patiño
Student's name: ____________________
Course Code: STA 145

Sections: A01 and A02.
Deadline: 6/7/2022, 5:00 p.m.
Total marks: 100

Dataset from: Kutner, M.H., C.J. Nachtsheim, J. Neter and W. Li (2005). Applied Linear Statistical Models, 5th ed. McGraw-Hill, New York.

The dataset (Demographic.txt - attached the description-variables of Data Set C.2) provides selected county demographic information with 17 variables for 440 of the most populous counties in the United States. Each line of the data set has an identification number with a county name and state abbreviation and provides information on 14 variables for a single county. Counties with missing data were deleted from the data set. The information generally pertains to the years 1990 and 1992. For each geographic region, regress the logarithm of the number of serious crimes ($\log(Y)$) against at least 5 predictor variables (you need to choose the predictors).

Q.1) Propose the exploratory analysis using the logarithm of the number of serious crimes ($\log(Y)$) and predictors. **Hint**: consider pairs(...), boxplot(...) and summary(...) in R (20 points).

Q.2) For the two selected models consider two different prior distributions:

- $$\boldsymbol{\beta} \sim \text{multivariate normal}(\boldsymbol{\beta}_0, \Sigma_0),$$
- $$\boldsymbol{\beta} \sim \text{multivariate normal}(\mathbf{0}, g(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^2),$$

with $\boldsymbol{\beta}_0 = \mathbf{0}$, $g = n$, $\Sigma_0 = 1000\boldsymbol{I}_p$ where $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix and where

$$\gamma = 1/\sigma^2 \sim gamma(\gamma; \nu_0/2, \nu_0\sigma_0^2/2),$$

with $\nu_0 = \sigma_0^2 = 1$.

1. Write the Monte Carlo and Gibbs sampler algorithms to sample from γ and $\boldsymbol{\beta}$. (20 points).
2. Compute 95% confidence intervals and 95% quantile based credible intervals for the parameters β_j, $j = 1, .., p$. (20 points).
3. Consider $\alpha = 0.01$ and compute the p-value for the two alternatives: (20 points)

$$H_0 : \beta_j = 0$$

versus

$$H_a : \beta_j \neq 0.$$

Compute the marginal posterior distribution of each β_j and compare the Bayesian and frequentist results to test:

$$H_0 : \beta_j = 0$$

versus

$$H_a : \beta_j \neq 0.$$

4. Obtain the residuals for each fitted model and prepare the diagnostic plots for each fitted model. State the conclusions. (20 points).