STAT 4640/7640 **Homework 5**

Due: March 15, 2022

- **Instructions**: Make sure your name is on your paper and your answers are clearly written.

1. Go through Listing 3.2 on page 79 in the text. Include your code and the figures. Try a few different choices of initial values for μ and σ^2 and compare your results.

2. Use the analysis of the NFL concussion data in Listing 3.3 of the text (and the course notes) as your guide to investigate the strikeout rate per game for Mizzou Softball. Over the course of seven seasons, 2015 through 2021, the number of strikeouts by Mizzou pitchers are given in the following table. Also given in the table is the number of games played per season.

Year	Strikeouts	Games
2015	343	58
2016	261	58
2017	271	57
2018	220	59
2019	278	60
2020	149	26
2021	317	59

   (a) Write out the three-stage hierarchical model like we did in class. Make sure to define your noninformative prior for γ.

   (b) Write a Gibbs sampling algorithm to obtain posterior inference for all model parameters.

   (c) Investigate the trace plots of all model parameters and comment on the convergence of the chain and burn-in.

   (d) Produce boxplots of the posterior distribution of the λ parameters for each year like in Figure 3.6 (left). Comment on any similarities or differences between the strikeout rate per game across the seven years.

   (e) Compute the posterior mean and 95% credible interval for each λ.

   (f) In 2020, the team played only 26 games. How does their strikeout rate per game compare in 2020 versus 2021 when they played 59 games? Provide statistical evidence to support your answer.

   (g) In each of the 2015 and 2016 season, Mizzou softball's record was 42-16. Therefore, the ratio of their win percentage for these two seasons is 1. Conduct a hypothesis test to determine whether or not the ratio of strikeout rates for these to seasons is greater than 1. Make sure to report the null and alternative hypotheses and the statistical evidence to support your conclustion.

3. Using the code from Listing 3.4 on page 88 in the text, conduct Bayesian linear regression with the `cars` dataset from Homework 3. Once again, distance is the response variable, and speed is the explanatory variable.

(a) Specify reasonable prior distributions for each parameter. Report what you choose.

(b) Obtain posterior draws for the parameters β_0, β_1, and σ^2. Make traceplots of your chain for each parameter and comment on convergence and burn-in.

(c) Make histograms of the posterior distribution for each parameter.

(d) Report the posterior mean and standard deviation of each parameter, as well as the 95% credible interval. Compare your estimates to the the least-squares estimates using `lm()`

(e) Create a scatter plot of the posterior chains of β_0 vs β_1 and comment on the correlation.

(f) (Extra Credit) Re-do the Bayesian linear regression without using a block update for β_0 and β_1. Specify your prior distributions. Derive the full conditional distributions of β_0 and β_1. Make sure to show your work. Write the Gibbs sampling algorithm in R using the univariate posteriors you derived and conduct posterior inference. Compare your results to those from the block-update Gibbs sampling algorithm.