Homework 5

Q1

The data in the file `enzyme.csv` gives the initial rate of reaction of an enzyme (y) and the substrate concentration (x). Consider the following nonlinear regression model:

$$y = \frac{\theta_1 x}{\theta_2 + x} + \epsilon$$

where $\epsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$, $\theta_1 > 0$, and $\theta_2 > 0$. Assume noninformative priors for θ_1, θ_2, and σ^2. You can choose appropriate prior distributions.

(a) Plot the marginal posterior densities of θ_1, θ_2, and σ^2. Use $\theta_1 = 200$, $\theta_2 = 0.1$, and $\sigma^2 = 100$ (equivalently, $\tau = 0.01$) for initializing the MCMC chain.

(b) Provide evidence that your model has converged, whether it is a trace plot, lack of divergences, the Gelman-Rubin statistic (Rhat), or something else.

(c) Compute 95% credible intervals, the mean, and the standard deviation for each of the three parameters. (From now on, we will rarely specify which type of credible interval—you may use the default for your chosen software.)

(d) Plot the posterior predictive distribution of y when $x = 0.75$ and provide the 95% credible intervals.

Q2

Walpole et al. (2007)[1] provide data from a study on the effect of magnesium ammonium phosphate on the height of chrysanthemums, which was conducted at George Mason University in order to determine a possible optimum level of fertilization, based on the enhanced vertical growth response of the chrysanthemums. Forty chrysanthemum seedlings were assigned to 4 groups, each containing 10 plants. Each was planted in a similar pot containing a uniform growth medium. An increasing concentration of $MgNH_4PO_4$, measured in grams per bushel, was added to each plant. The 4 groups of plants were grown under uniform conditions in a greenhouse for a period of 4 weeks. The treatments and the respective changes in heights, measured in centimeters, are given in the following table:

[1] Walpole, W. A., Myers, R. H., Myers, S. L., and Ye (2007). *Probability and Statistics for Engineers and Scientists* (9th ed.). Pearson.

Treatment			
50 g/bu	100 g/bu	200 g/bu	400 g/bu
13.2	16.0	7.8	21.0
12.4	12.6	14.4	14.8
12.8	14.8	20.0	19.1
17.2	13.0	15.8	15.8
13.0	14.0	17.0	18.0
14.0	23.6	27.0	26.0
14.2	14.0	19.6	21.1
21.6	17.0	18.0	22.0
15.0	22.2	20.2	25.0
20.0	24.4	23.2	18.2

Solve the problem as a one-way ANOVA. Use STZ constraints on treatment effects.

(a) Do different concentrations of $MgNH_4PO_4$ affect the average attained height of chrysanthemums? Look at the 95% credible sets for the differences between treatment effects.

(b) Find the 95% credible set for the contrast $\mu_1 - \mu_2 - \mu_3 + \mu_4$.

(c) In a standard one-way ANOVA, we assume constant variance σ^2 for each group. If you relax that assumption and put a prior on each group's standard deviation (σ_i for $i = 1, \ldots, 4$), do the results from (a) and (b) change? Do the contrasts between the posterior distributions of each σ_i show that they were significantly different?

Q3

The data set (available as `wolves.csv`) described below provides skull morphometric measurements on wolves (*Canis lupus L.*) coming from two geographic locations: Rocky Mountain (0) and Arctic (1). The original source of the data is from Jolicoeur (1959)[2], and many authors have subsequently used this data to illustrate various multivariate statistical procedures.

The goal of Jolicoeur's study was to determine how location and gender affect skull shape among wolf populations. There were 9 predictor variables measured (see Table 1).

[2] Jolicoeur, P. (1959). Multivariate geographical variation in the wolf *Canis lupus L. Evolution*, **13**(3), 283–299. Data here are given in inches.

Table 1: Wolf skull morphometric data (in inches) from Jolicoeur (1959)[3].

Variable	**Description**
location	0 = Rocky Mountain, 1 = Arctic
gender	0 = male, 1 = female
x_1	Palatal length
x_2	Postpalatal length
x_3	Zygomatic width
x_4	Palatal width (outside first upper molars)
x_5	Palatal width (inside second upper molars)
x_6	Width between postglenoid foramina
x_7	Interorbital width
x_8	Least width of the braincase
x_9	Crown length of the first upper molar

(a) Try a frequentist logistic regression on the data (in Python, you can use the statsmodels package or sklearn). What are the results?

(b) Set up a Bayesian logistic regression. Try at least three separate models, changing regression coefficient variance to increasingly informative values for each. What do you observe in the results? How do they differ from the frequentist model and from each other?

(c) Re-sample the model with only three predictors: gender, x_3, and x_7. Give an estimate and credible interval of the probability that a female wolf with measures $x_3 = 5.28$ and $x_7 = 1.78$ comes from an Arctic habitat.