STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN DATA SCIENCE & ANALYTICS
CAT 2- **Open Book**
DSA 8505: Bayesian Statistics

DATE: 21st Mar 2025

Instruction

(a) Answer All Question

(b) Scan and submit your answer sheet through the Google Classroom by 23h59, 27th March 2025.

1. Suppose we are modeling the probability of a patient recovering from a disease ($Y_i = 1$) based on the number of days they adhered to a prescribed treatment (X_{1i}) and their age in years (X_{2i}). We assume a logistic regression model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

Given the following dataset:

Days on Treatment (X_{1i})	**Age** (X_{2i})	**Recovered** (Y_i)
1	25	0
2	30	0
3	35	1
4	28	1
5	40	1
2	45	0
6	50	1
3	33	1
4	27	1
5	29	1

We assume the following priors:

$$\beta_0 \sim \mathcal{N}(0, 10)$$
$$\beta_1 \sim \mathcal{N}(0, 10)$$
$$\beta_2 \sim \mathcal{N}(0, 10)$$

The likelihood function is given by:

$$P(Y|\beta) = \prod_{i=1}^{n} p_i^{Y_i}(1 - p_i)^{1-Y_i}$$

Using the MCMC method in Python, we estimated the posterior distributions of β_0, β_1, and β_2 as follows:

- β_0 has a mean of -8 with a 94% HDI of (-18, 2).
- β_1 has a mean of 3.8 with a 94% HDI of (0.5, 8.0).
- β_2 has a mean of -0.2 with a 94% HDI of (-1.5, 1.0).

Compute the probability of recovery if a patient follows the treatment for 3 days and is 30 years old, and for a patient who follows the treatment for 5 days and is 35 years old. Interpret the results.

2. Bayesian logistic regression is used to model binary outcomes, such as predicting whether a patient has a disease based on their age and cholesterol level. Here is a sample dataset:

Disease (1 = Yes, 0 = No)	**Age (years)**	**Cholesterol (mg/dL)**
0	45	180
1	52	220
1	60	250
0	35	160
0	40	170
1	55	240
0	30	150
1	65	270
0	42	175
1	58	230

(a) Explain the role of prior distributions in Bayesian logistic regression and discuss their impact on inference.

(b) The logistic function (sigmoid) is used to transform the linear combination of predictors into a probability. Explain why it is appropriate for modeling binary outcomes.

(c) In the Bayesian model given, the posterior estimates of beta's are:

$$\beta_0 = -0.1, \quad \beta_{\text{age}} = -0.25, \quad \text{and} \quad \beta_{\text{cholesterol}} = 0.018.$$

Compute the probability of disease for a 50-year-old patient with a cholesterol level of 200 mg/dL. Interpret the result in the context of disease diagnosis.