Strathmore
UNIVERSITY

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

DATE: 17 Feb 2026

Instruction

(a) Answer All Question

(b) Scan and submit your answer sheet through the e-learning by 21h59, 17th Feb 2026.

---

1. A security agency uses an automated facial recognition system to identify individuals on a watchlist at a busy airport. Based on historical records, it is known that only 0.5% of travelers are actually on the watchlist.

   The facial recognition system has the following performance characteristics:

   - If a traveler is on the watchlist, the system correctly flags them with probability 95% (true positive rate).
   - If a traveler is not on the watchlist, the system incorrectly flags them with probability 3% (false positive rate).

   A randomly selected traveler is scanned by the system and is flagged as a match.

   Required:

   (a) Define the prior probability that a randomly selected traveler is on the watchlist. [**2 mk**]

   (b) Define the likelihood of the system flagging a traveler, conditional on whether the traveler is on the watchlist or not. [**4 mk**]

   (c) Compute the marginal probability that a randomly selected traveler is flagged by the system. [**4 mk**]

   (d) Using Bayes' Theorem, compute the posterior probability that the traveler is actually on the watchlist given that they were flagged by the system. [**5 mk**]

(e) Interpret the result from a policy perspective: Based on the computed posterior probability, discuss whether airport security should immediately detain the traveler or apply additional screening procedures. **[5 mk]**

2. Consider the problem of modelling monthly electricity consumption y (in kWh) using the average monthly temperature x (in degrees Celsius). Two competing regression models are proposed:

Model 1: Linear Temperature Effect

$$M_1: \quad y = \beta_0 + \beta_1 x + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Model 2: Quadratic Temperature Effect

$$M_2: \quad y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

where $\beta_0, \beta_1, \beta_2$ are unknown regression parameters and σ^2 is the error variance.

The observed dataset is shown in Table 1.

Average Temperature x	Electricity Use y
10	220
15	240
20	280
25	350
30	460

Table 1: Observed electricity consumption data

Assume equal prior probabilities for the two competing models:

$$P(M_1) = P(M_2) = 0.5.$$

The prior distributions for the parameters are specified as:

$$\beta_j \sim \mathcal{N}(0, 10^2), \quad j = 0, 1, 2,$$

$$\sigma^2 \sim \text{Inverse-Gamma}(1, 1).$$

(a) Using the ordinary least squares (OLS) method, estimate the regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ for Model M_1. **[5 mk]**

(b) Write down the likelihood function for Model M_1. **[2 mk]**

(c) Using Bayes' theorem, write the posterior distribution of the regression parameters under Model M_2. **[1 mk]**

(d) Explain why Markov Chain Monte Carlo (MCMC) methods are required to obtain samples from the posterior distribution. **[2 mk]**

(e) The marginal likelihoods for the two models are estimated numerically as:

$$P(D \mid M_1) = 0.015, \qquad P(D \mid M_2) = 0.045.$$

(i) Compute the Bayes Factor in favour of Model M_2. **[1 mk]**

(ii) Calculate the posterior model probabilities $P(M_1 \mid D)$ and $P(M_2 \mid D)$. **[2 mk]**

(iii) Interpret the results of the Bayesian model comparison. **[1 mk]**

(f) Suppose the posterior mean estimates of the regression coefficients are obtained as:

- Model M_1: $\hat{\beta}_0 = 150$, $\hat{\beta}_1 = 8.5$
- Model M_2: $\hat{\beta}_0 = 120$, $\hat{\beta}_1 = 6.2$, $\hat{\beta}_2 = 0.45$

(i) Write down the general Bayesian Model Averaging (BMA) formula for regression coefficients. **[2 mk]**

(ii) Compute the BMA estimates for β_0, β_1, and β_2. **[3 mk]**

(iii) Write the final BMA regression equation. **[1 mk]**