United States International University-Africa

Education to take you places

SCHOOL OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF DATA SCIENCE AND ANALYTICS

FALL 2024 – QUIZ 4

COURSE CODE: STA 4030A

UNIT NAME: BAYESIAN INFERENCE ND DECISION THEORY

DATE: 08TH NOVEMBER 2024 **TOTAL MARKS**: 30 MARKS

INSTRUCTIONS:

For this exercise:

1. ANSWER ALL QUESTIONS
2. Do all your working in the Rmarkdown (.rmd).
3. Submissions should be in a **`.rmd` file**
4. NO SUBMISSIONS SHOULD BE DONE VIA EMAIL

QUESTIONS:

1. Concentrations of the pollutants aldrin and hexachlorobenzene (HCB) in nanograms per liter were measured in ten surface water samples, ten mid-depth water samples, and ten bottom samples from the Wolf River in Tennessee. The samples were taken downstream from an abandoned dump site previously used by the pesticide industry. The full data set can be found at this URL. For this problem, we consider only HCB measurements taken at the bottom and the surface. The question of interest is whether the distribution of HCB concentration depends on the depth at which the measurement was taken. The data for this problem are given below.

   Data:

Surface	Bottom
3.74	5.44
4.61	6.88
4.00	5.37
4.67	5.44
4.87	5.03
5.12	6.48
4.52	3.89
5.29	5.85
5.74	6.85
5.48	7.16

   Assume the observations are independent normal random variables with unknown depth-specific means θ_s and θ_b and precisions ρ_s and ρ_b. Assume independent improper reference priors for the surface and bottom parameters:

$$g(\theta_s, \theta_b, \rho_s, \rho_b) = g(\theta_s, \rho_s)g(\theta_b, \rho_b) \propto \rho_s^{-1}\rho_b^{-1}.$$

   This prior can be treated as the product of two normal-gamma priors with $\mu_s = \mu_b = 0$, $k_s = k_b = 0$, $a_s = a_b = -\frac{1}{2}$, and $b_s = b_b = \infty$. (These are not valid normal-gamma distributions, but you can use the usual Bayesian conjugate updating rule to find the posterior distribution.) Find the joint posterior distribution for the parameters $(\theta_s, \theta_b, \rho_s, \rho_b)$. State the type of distribution and the posterior hyperparameters. Find 90% posterior credible intervals for θ_s, θ_b, ρ_s, and ρ_b. Comment on your results.

2. Use direct Monte Carlo to sample 10,000 observations from the joint posterior distribution of $(\theta_s, \theta_b, \rho_s, \rho_b)$. Use your Monte Carlo samples to estimate 90% posterior credible intervals for θ_s, θ_b, ρ_s, and ρ_b.
3. Use your direct Monte Carlo sample to estimate the probability that the mean bottom concentration θ_b is higher than the mean surface concentration θ_s and to estimate the probability that the standard deviation σ_b of the bottom concentrations is higher than the standard deviation σ_s of the surface concentrations.
4. Comment on your analysis. What are your conclusions about the distributions of surface and bottom concentrations? Is the assumption of normality reasonable? Are the means different for surface and bottom? The standard deviations?