

INSTRUCTIONS:

- Answer all questions in a single Word document. Please convert to .pdf at the end.
- Label questions clearly, as it is done on this question paper.
- All results accurate to 2 decimal places.
- Show all derivations, formulas, code, sources and reasoning.
- Intervals should cover 95% probability unless stated otherwise.
- No communication software, devices or websites may be accessed prior to submission.

Question 1

This entire test is based on a supplied data set given on your LMS as ‘Bayes2019Test2regData.csv’.

The target variable of interest is the Forced Exhalation Volume (FEV) of a sample of children of various ages. It is a proxy for lung capacity. In this question we will analyse FEV on the natural log scale with a Normal distribution.

Explanatory variables included in this data set include the age of the child, their height, their gender, and whether or not they smoke.

- (a) Read in the data. Create a new variable $\ln\text{FEV}$ and show that it has a mean close to 1. [3]
- (b) Do an ordinary linear regression of $\ln\text{FEV}$ on age, height, gender, and smoking status. Mention which explanatory variables are deemed significant. [3]
- (c) Create a function that generates random initial values for each model parameter according to the estimates and standard errors reported by the regression above. [2]
- (d) Repeat the regression above by simulating at least 10000 samples from the joint posterior of the parameters, assuming independent, vague, but proper priors for all parameters. Give the posterior mean estimates of all the parameters. [7]
- (e) Use your simulations to calculate a Bayesian equivalent of p-values for each coefficient (to attempt to see how they differ from 0) and discuss any differences with the ordinary model. [4]

- (f) Simulate a sample of size at least 10000 from the posterior predictive distribution of every observation. Reverse the log transform and calculate 95% HPD intervals for every observed FEV value. [5]
- (g) Calculate the empirical coverage of your fitted model. Is it in line with what you expected? [2]
- (h) Report the DIC of your model. Now modify your model to allow for a variance that changes linearly with the age of the child. Give the new DIC and state whether it is an improvement. [7]

Total for Question 1: 33

Question 2

In this question we want to model the FEV as given with a Gamma distribution. We would like the expected value to change with age and gender only, and we want the variance to be assumed constant throughout.

This model requires a parameter transformation and priors as follows:

$$\begin{aligned}
 FEV_i &\sim \text{Gamma}(\alpha, \lambda) \\
 \alpha &= \mu^2 \tau \\
 \lambda &= \mu \tau \\
 \mu &= \exp(\beta_0 + \beta_1 \text{age} + \beta_2 \text{gender}) \\
 \beta_i &\sim N(0, 10^2) \quad \forall i \\
 \tau &\sim \text{Exp}(0.1)
 \end{aligned}$$

- (a) Simulate a sample from the joint posterior distribution (at least 10000 draws) and give Bayesian a equivalent of p-values for β_0, β_1 , and β_2 (to attempt to see how they differ from 0). [7]
- (b) Simulate a sample of size at least 10000 from the posterior predictive distribution of every observation. Give 95% HPD intervals for every observed FEV value. Comment on the coverage. [5]

Total for Question 2: 12