Homework 1

Data

The data in '`MedicalData.csv`' represent health records from a sample of 155 total individuals of varying ages from a broader population. The data set contains several variables. For this homework assignment, we seek to learn the probability that mean (μ) body mass index (BMI) in the population of 'young adults' is greater than 30 (i.e., obese). The 'young adult' subsample consists of $n = 29$ observations.

Questions

Prepare a written response to the following, using Overleaf. The assignment shouldn't be longer than 10 (double-spaced, excluding title page, references, and appendices). Due Tues., Jan. 24, at the beginning of the class period. Please submit the assignment as a PDF through CANVAS.

1. Derive the full-conditional distribution for $[\sigma^2|\mathbf{y}, \mu]$ associated with the Normal-Normal-Inverseχ^2 model. You will need this for Gibbs sampling.

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu, \sigma^2)\ , i = 1, \ldots, n\ , \\
\mu &\sim \mathrm{N}(\mu_0, \sigma_0^2)\ , \\
\sigma^2 &\sim \text{Inverse}\chi^2(\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-(\nu/2+1)} \exp(-1/(2\sigma^2))\ .
\end{aligned}
$$

2. Construct and use an MCMC algorithm to fit a Normal-Normal-Inverseχ^2 model to the BMI data to answer the question described in the "Data" section above. Briefly describe the procedure and include the R code as an appendix. Concisely describe the model, including prior specifications, the algorithm, the results (including the number of MCMC iterations, visual inspection of trace plot, burn-in period if necessary, and comment on mixing).

3. Provide a graphic displaying the marginal posterior distributions for μ and σ^2 resulting from the model fit to the data. Report the posterior mean and 95% equal-tailed credible interval for both μ and σ^2 in a table.

4. Use the MCMC output to calculate the posterior probability that the mean BMI for young adults is greater than 30 (i.e., $\mathrm{P}(\mu > 30|\mathbf{y})$). What can you infer as a result of this analysis?

5. Discuss the assumptions of the model relative to these data being analyzed and whether or not you feel they were met.