Bayesian Inference for Surveys

Homework # 4

Due: February 13, 2017

A simple random sample of 10 villages was selected from a population of 34 villages in a subdivision of one of the States in India. From the administrative records and the satellite imagery analysis, the total cultivated area (in acres) in these 34 villages were known and are given in the Excel spreadsheet (*hwdata.xlsx*). The field workers visited the sampled villages and obtained the area (in acres) used for producing wheat. The primary goal is to estimate the total area used for wheat production in the population of 34 villages. The Excel spreadsheet also provides area under wheat for the 10 sampled villages.

1. For now, ignore the total cultivated area. What is the posterior distribution of the total area under wheat? Compute its posterior mean, standard deviation and a 95 % credible interval. You may use a normal distribution for the area under wheat and a non-informative prior for all the parameters.

2. Consider using the total cultivated area as an auxiliary variable. Develop a regression model predicting area under wheat using the total cultivated area as a predictor. Assume that the residuals are normally distributed and all parameters have non-informative prior distribution. Using a missing data package, or otherwise, construct the posterior predictive distribution of the population total of area under wheat. Compute its posterior mean, standard deviation and 95% highest posterior credibility interval.

3. Compare the inferences obtained in (1) and (2) and comment on the usefulness of the administrative data.