The data set LLFS.subset.csv includes data about few genetic polymorphisms that could be associated with longevity (outcome = 1)

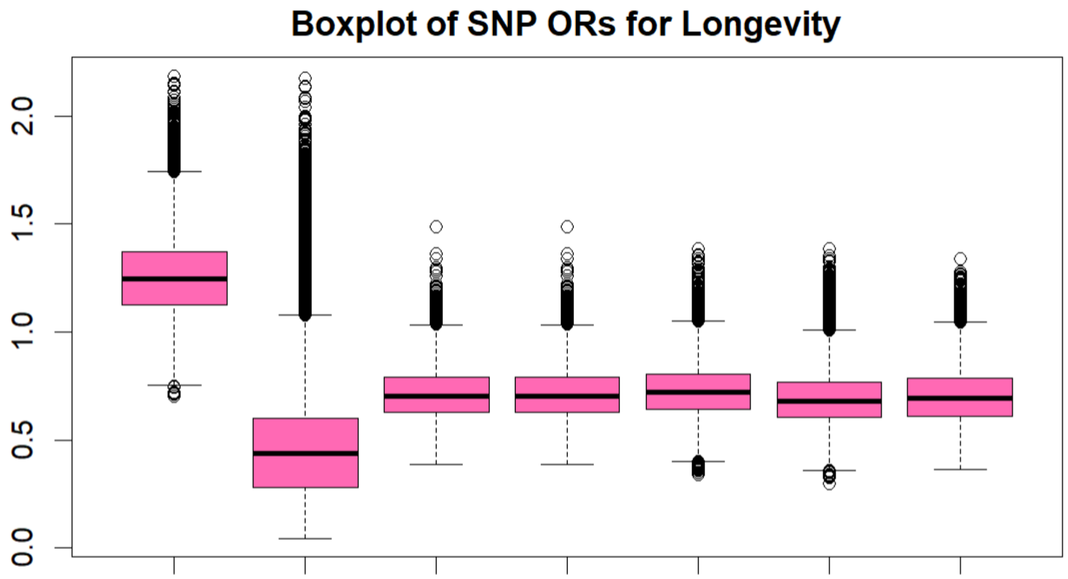
Read the data set and create a new index variable that represent family membership in the data. How many families in the study have only one individual?
228 out of 656 families have only one individual.

Implement a Bayesian model that describes the genetic association between each SNP and the outcome using logistic regression adjusted for 2 PCs, and sex. Use a hierarchical model with random intercept to accommodate the within-family relations. Run the Bayesian analysis of the association between longevity and each of the 7 SNPs using 10,000 iterations of the MCMC. Specify if you run the analysis on the SCC or on your computer and provide details of the amount of time taken by the analysis.
The generation and analysis of 10,000 Monte-Carlo samples for all seven SNPs took a total of 16 minutes and 53 seconds on my very old Dell laptop.

Produce a summary table with the output that includes the estimate of the genetic effect of each SNP (genetic effect = OR for longevity for carrier of one coded allele) and 95% credible intervals, and boxplots of the posterior distributions. Which SNPs are significantly associated with longevity?

SNP	Median Parameter	95% Credible Interval	Median OR	95% Credible Interval
SNP1	0.2189	-0.06709, 0.5066	1.2447	0.93511, 1.6597
SNP2	-0.8298	-2.58211, 0.2917	0.4362	0.07561, 1.3387
SNP3	0.2150	-0.06443, 0.5057	1.2400	0.93760, 1.6582
SNP4	-0.3513	-0.6847, -0.02023	0.7038	0.5043, 0.97998
SNP5	-0.3285	-0.6808, 0.00257	0.7200	0.5062, 1.00257
SNP6	-0.3866	-0.7373 , -0.03057	0.6794	0.4784, 0.96990
SNP7	-0.3663	-0.7429, -0.01998	0.6933	0.4757, 0.98021

SNP4, SNP6, and SNP7 are associated with longevity, as their 95% Credible Intervals for the parameter don't include 0 and for the OR don't include 1.



Use the Geweke statistics to check whether the analyses of all 7 SNPs have converged.

SNP	First Arm Z-Score	Second Arm Z-Score	Third Arm Z-Score
SNP1	-0.1261	0.3992	4.0290
SNP2	1.3960	-0.0832	1.1310
SNP3	1.7260	0.3584	0.0113
SNP4	0.3216	0.5676	0.8880
SNP5	-0.8261	-0.3046	0.1872
SNP6	1.4320	-0.4334	1.3810
SNP7	0.9483	1.3290	-2.2240

Geweke z-statistics comparing the equality of means between the first 10% and last 50% of the Monte Carlo simulations are listed above. All the SNPs displayed convergence except for SNP1 and SNP7.

Compute the Gelman-Rubin statistics using 3 chains for SNP 2. Discuss your findings.

The scale reduction factor estimate for the parameter of SNP2 was 1.63 and the upper confidence interval was 2.68. That number is well above 1, which indicates nonconvergence due to large differences in the between-chain and within-chain variances.

Compute the Gelman-Rubin statistics for SNP 5. Discuss your findings.

The scale reduction factor estimate for the parameter of SNP2 was 1 and the upper confidence interval was also 1. That indicates convergence between the three chains and extremely small differences in the between-chain and within-chain variances.