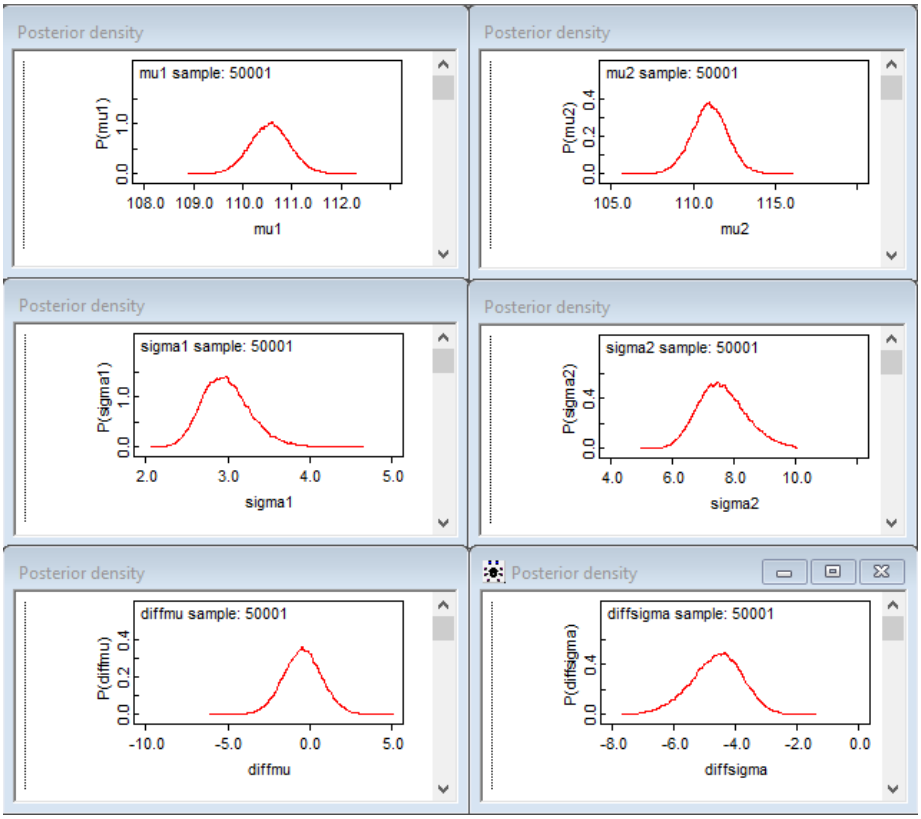
AC50001 ASSIGNMENT ON BAYESIAN INFERENCE USING OPENBUGS

JULIA GRATSOVA 130017011

Part 1: Spot the difference

Use OpenBUGS to model each of the two given data sets as Gaussian and run MCMC to infer all the parameters of these two Gaussians. In addition, set up variables for (i) the difference between the two means and (ii) the difference between the two standard deviations.

POSTERIOR DENSITY PLOTS



NODE STATS

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu1	110.6	0.4045	0.001766	109.8	110.6	111.3	50000	50001

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu2	111.0	1.1	0.004919	108.8	111.0	113.1	50000	50001

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
sigma1	2.985	0.2958	0.001454	2.477	2.962	3.635	50000	50001

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
sigma2	7.616	0.7795	0.00378	6.239	7.558	9.305	50000	50001

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
diffmu	-0.4405	1.172	0.005425	-2.726	-0.4415	1.879	50000	50001

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
diffsigma	-4.631	0.8343	0.003989	-6.402	-4.583	-3.107	50000	50001

Conclusion

1 MCMC chain was created for each variable of interest with a sample set size 100,000 while discarding the first 50,000 results.

Group 1:

n1 = 55

$\bar{X}_1$= 110.6

s1 = 0.4045

Group 2:

n2 = 48

$\bar{X}_2$ = 111.0

s2 = 1.1

Check the assumption of the variance equality $s_1^2 = s_2^2$

s_1^2/s_2^2 = 0.135

As the ratio is less than 0.5 it is unlikely that they are different.

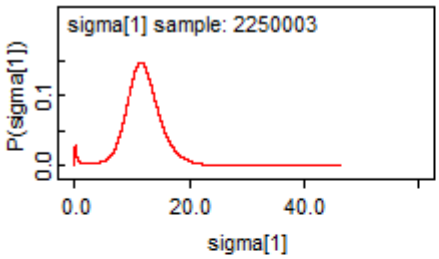
The diffmu Gaussian posterior plot shows the posterior probability below and above zero which makes it impossible to tell with determination whether one is greater than another.

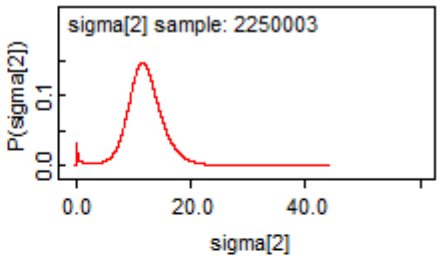
The diffsigma Gaussian posterior plot shows that all of the posterior probability is below zero, which allows to say that sigma2 is larger than sigma1.

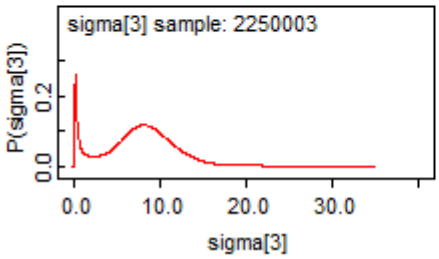
Part 2: Multiple observers

Using the Seven Scientists model from Lab Sheet 2 as a starting point, specify a suitable model for this situation. The model should enable inference of the N quantities of interest as well as variables indicative of the skill level of each of the M observers.

POSTERIOR DENSITY PLOTS



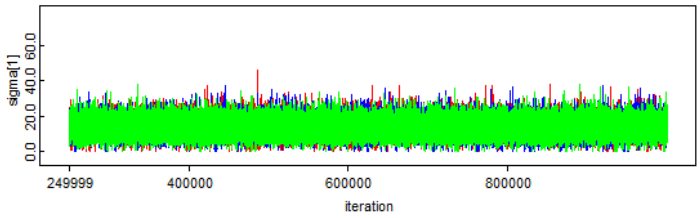


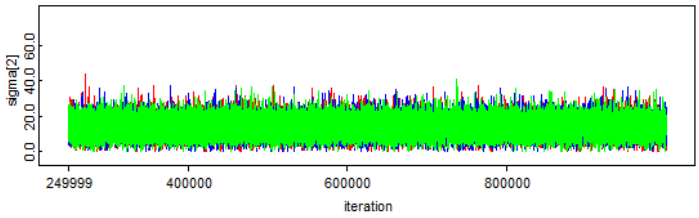


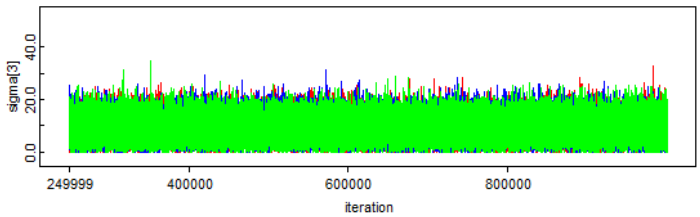
NODE STATISTICS

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
sigma[1]	11.86	3.387	0.02313	4.561	11.8	18.77	250000	2250003
sigma[2]	11.87	3.392	0.02381	4.511	11.81	18.77	250000	2250003
sigma[3]	7.31	3.979	0.03768	0.07584	7.677	14.7	250000	2250003

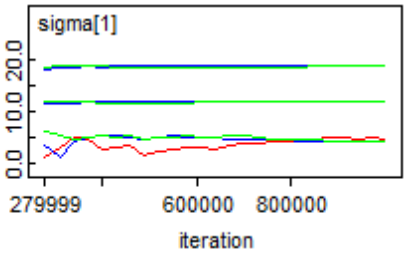
HISTORY

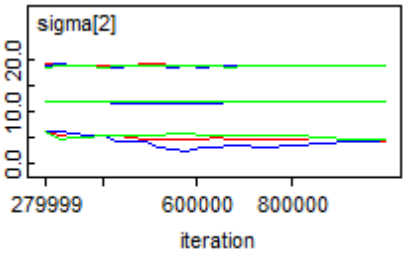


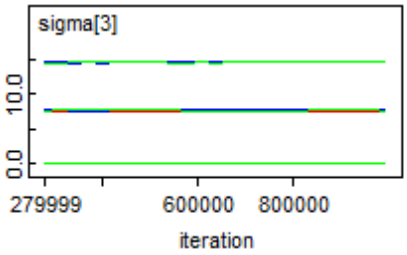




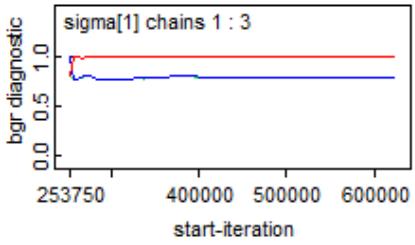
RUNNING QUANTILES

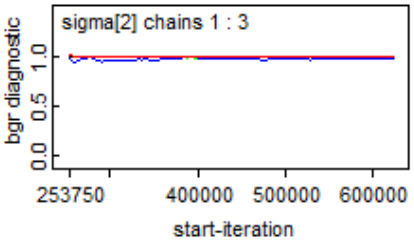


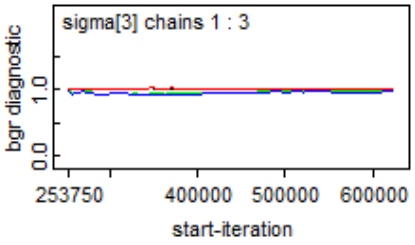




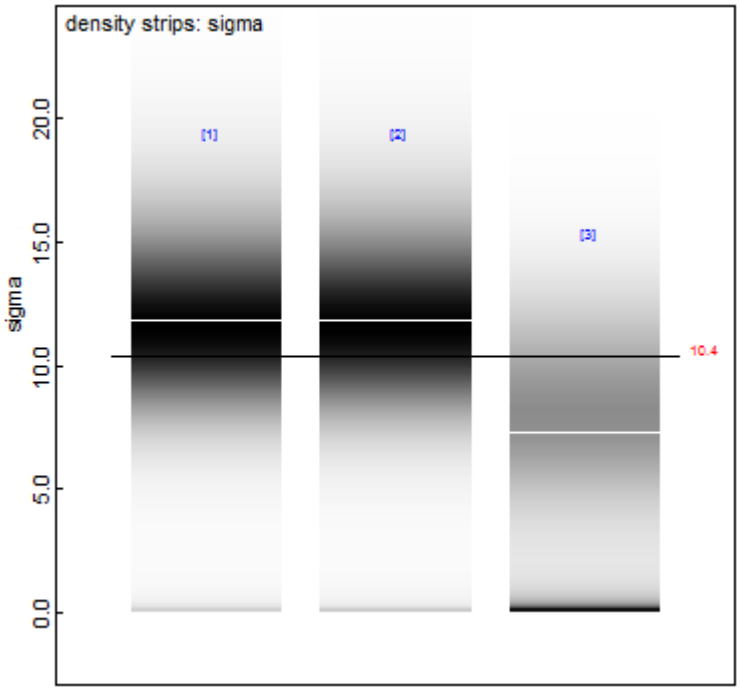
GELMAN-RUBIN CONVERGENCE STATISTIC

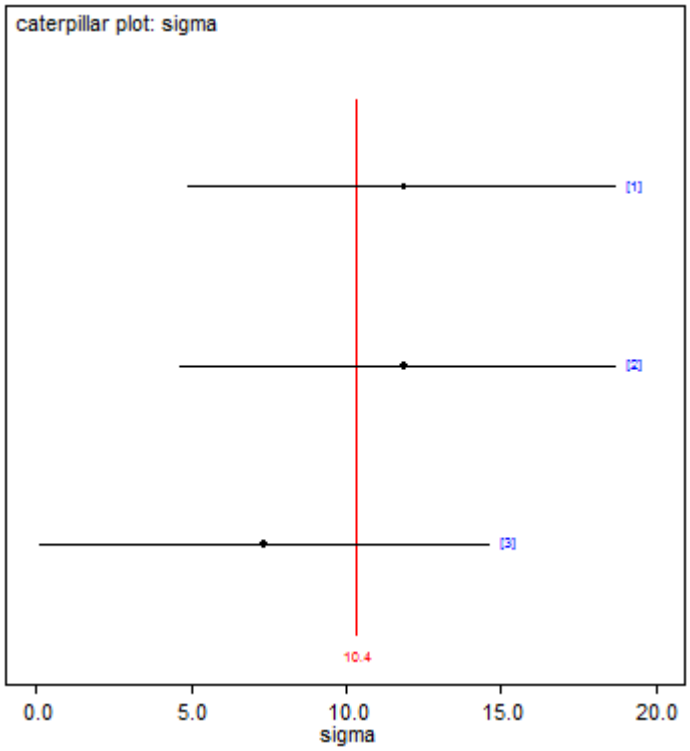






DENSITY STRIPS AND CATERPILLAR PLOT





Conclusion

Before using the samples for parameter estimation , I run 3 chains in parallel and discarded the first portion of the results. I then generated the bgr diagnostic chart from within the Sample Monitor Tool as it is designed to calculate the Gelman-Rubin statistic, the general rule of which is that each parameter is < 1.10. The produced chart satisfies the rule and it can be said that the model converged.

The density strips and caterpillar plots show that the average mean is 10.4 and both sigma1 and sigma2 are above the average mean. Looking at the results produced, it can be seen that sigma [3] is the lowest, therefore the observer 3 produced the most accurate result out of the three.

Part 3

While the time taken to read and research the extra material is hard to determine, the completion of the assignment itself took approximately 17 hours to complete.