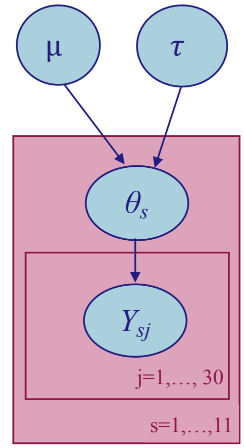
***Problem 1*:**

Section 18.4 of Gelman, et al. (2013) analyzes data on reaction times for 11 non-schizophrenic and 6 schizophrenic subjects. The data set can be found at:

http://www.stat.columbia.edu/~gelman/book/data/schiz.asc

The first 11 rows are data for the non-schizophrenic subjects. (We analyzed data from one of these subjects in Unit 5.) Consider a hierarchical model for the natural logarithms of reaction times of the non-schizophrenic subjects. A plate diagram for the model is shown at the right.



- Following Gelman, et al., assume that the logarithms of the response times are independent normal random variables with person-specific mean θ_s (s = 1,…11).
- Following Gelman, et al., assume all the observations have the same standard deviation σ. Assume σ is known and equal to the average of the 11 sample standard deviations.
- The 11 means θ_s, s=1, …, 11, are independent and identically distributed normal random variables with mean μ and standard deviation τ.
- The parameters μ and τ are independent of each other.
- The unknown mean μ has a normal distribution with mean 5.52 and standard deviation 0.22. This reflects a prior 95% credible interval of [162, 385] ms for the population average reaction time, which is consistent with the literature on reaction times.
- The inverse variance $1/\tau^2$ of the θ_s has a gamma distribution with shape ½ and scale 50. This reflects weak prior information focused on a value of 25 for $1/\tau^2$, or 0.2 for τ. This is consistent with typical variability of reaction times.

Using the formulas in the Unit 5 notes and the posted examples as a guide, find the posterior distribution for each of the unknown parameters given the other parameters. *Remember that the hyperparameters of the posterior distribution will be formulas involving the other hyperparameters.*

***Solution*:**

This hierarchical model is structurally almost the same as the math test scores example in Module 7. The main difference is we are assuming a known standard deviation σ=0.157 (the average of the 11 sample standard deviations), whereas in the math test scores problem, we assumed an unknown precision with a gamma prior distribution. Therefore, the posterior distributions can be found by using the formulas on Page 45 of the Unit 7 notes and adjusting according to the differences between them:

- Given ($\underline{\theta}$, τ), μ is normally distributed with
  - mean $\frac{11\bar{\theta}/\tau^2+5.52/0.22^2}{11/\tau^2+1/0.22^2} = \frac{\sum_s \theta_s/\tau^2+5.52/0.22^2}{11/\tau^2+1/0.22^2} = \frac{\frac{\sum_s \theta_s}{\tau^2}+114.08}{11/\tau^2+20.66}$
  - standard deviation $(11/\tau^2+20.66)^{-1/2}$
- Given ($\underline{\theta}$, μ), $1/\tau^2$ has a gamma distribution with
  - shape ½+11/2 =6
  - scale $\left(1/50+\frac{1}{2}\left(\sum(\theta_s-\mu)^2\right)\right)^{-1}$
- Given (μ, τ, σ, $\underline{y}$), each θ_s is normally distributed with
  - mean $\frac{\sum_{i=1}^{n_s} y_{si}/\sigma^2+\mu/\tau^2}{30/\sigma^2+1/\tau^2} = \frac{\sum_{i=1}^{30} y_{si}/\sigma^2+\mu/\tau^2}{1217.5+1/\tau^2}$
  - standard deviation $(30/\sigma^2+1/\tau^2)^{-1/2} = (1217.5+1/\tau^2)^{-1/2}$

Notice that the hyperparameters for each of these distributions are given as ***formulas*** not numbers. It is a common error to provide numbers (for example, saying that μ has a normal distribution with mean 5.7 and standard deviation 0.04). This is incorrect. We cannot give a number for the mean and standard deviation of μ because they depend on the specific values of $\underline{\theta}$ and τ, which will change every iteration of Gibbs sampling.

Problem 2:

Use Gibbs sampling to draw 5000 samples from the posterior distribution of the parameters μ, τ, and θ_s, s=1, …, 11. Find 95% posterior credible intervals for each of these parameters. You may use JAGS or you may directly implement a Gibbs sampler. Use the examples provided as a guide.

Solution:

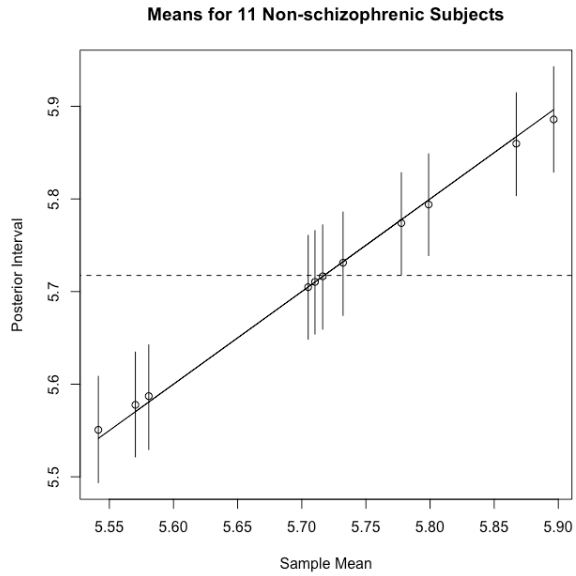
This can be done "by hand" or with JAGS. For these solutions, we code the Gibbs sampling "by hand." Using the formulas from Problem 1, Gibbs sampling proceeds as follows:

1. Initialize $\mu^{(1)}$, $\tau^{(1)}$, and $\theta_s^{(1)}$.
2. For $k = 2, \ldots, 5000$
    a. Sample $\mu^{(k)}$ from a normal distribution with mean $\frac{\sum_s \theta_s/\left(\tau^{(k-1)}\right)^2 + 5.52/0.22^2}{11/\left(\tau^{(k-1)}\right)^2 + 1/0.22^2}$ and standard deviation $(11/\tau^2 + 1/0.22^2)^{-1/2}$
    b. Sample a value u from a gamma distribution with shape 6 and scale $\left(1/50 + \frac{1}{2}\left(\sum\left(\theta_s{}^{(k-1)} - \mu^{(k)}\right)^2\right)\right)^{-1}$, and set $\tau^{(k)} = 1/u^{1/2}$.
    c. For s =1, …, 11, sample $\theta_s^{(k)}$ from a normal distribution with mean $\frac{\sum_{i=1}^{30} y_{si}/\left(\sigma^{(k)}\right)^2 + \mu^{(k)}/\left(\tau^{(k)}\right)^2}{30/\sigma^2 + 1/\left(\tau^{(k)}\right)^2}$ and standard deviation $\left(30/\sigma^2 + 1/\left(\tau^{(k)}\right)^2\right)^{-1/2}$.

R code for the Gibbs sampling is provided on Blackboard. You should compare the formulas given in these solutions with the R code to make sure you understand how the sampling works.

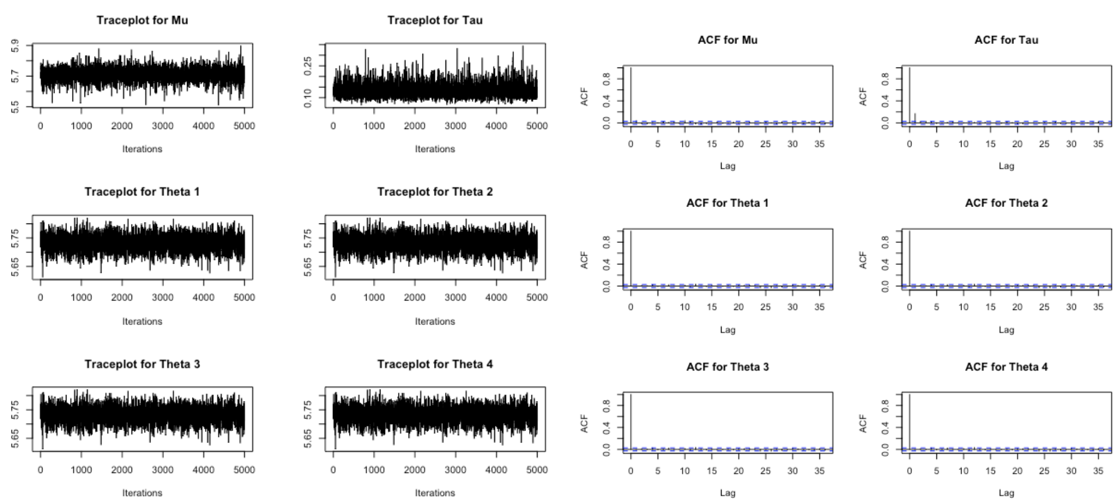
The table shows lower and upper ends of 95% credible intervals for the eleven subject means, and for the parameters μ, τ, and σ. The caterpillar plot gives a visual comparison of the intervals. It is very clear from the plot that the means for different subjects are different.

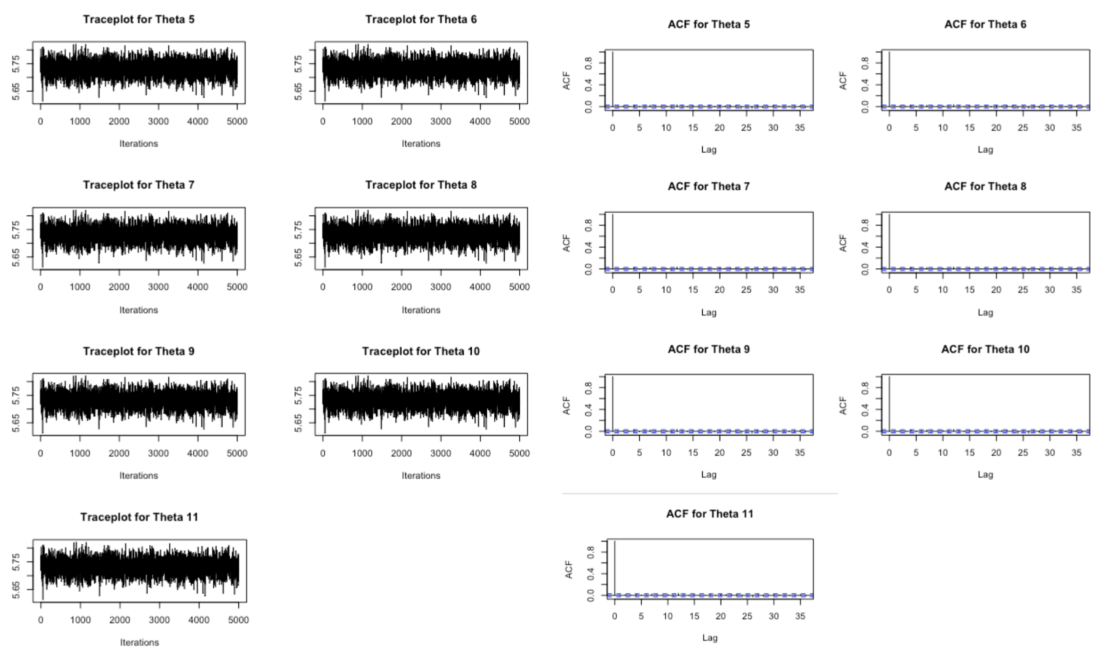
Subject	2.5%	97.5%
θ_1	5.676	5.784
θ_2	5.832	5.943
θ_3	5.655	5.764
θ_4	5.652	5.76
θ_5	5.524	5.634
θ_6	5.739	5.848
θ_7	5.806	5.915
θ_8	5.532	5.642
θ_9	5.494	5.605
θ_{10}	5.721	5.829
θ_{11}	5.661	5.771
μ	5.626	5.793
τ	0.087	0.212



One common error is to estimate eleven different intervals for μ and τ. There is a separate mean θ_s for each subject, but there is only one of each of the hyperparameters μ, τ, and σ.

It is useful to examine some MCMC diagnostics for this problem. Traceplots for all the parameters look stationary; there doesn't seem to be a need for a burnin. Autocorrelation plots for μ, σ, and θ_s, s=1, …, 11 show almost no autocorrelation. Only τ shows a small amount of first-order autocorrelation, but it is nearly zero by second order.





In line with the autocorrelation plots, the effective sample sizes are very high for all parameters except τ, and the effective sample size for τ is over 3500, indicating we have accurate estimates for all parameters. Every run these values will be slightly different.

μ	τ	θ_1	θ_2	θ_3	θ_4	θ_5
4607	3570	5000	5000	5000	4825	4512

θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
5000	4708	4543	4668	4583	4579

Problem 3:

Consider the following empirical Bayes model for the subject-specific means θ_s, s=1, …, 11.

- Assume the observation standard deviation σ is known and equal to the average of the 11 sample standard deviations of the log reaction times of the 11 subjects.
- Assume the means θ_s, s=1, …, 11 are normally distributed with mean μ equal to the grand mean of all the log reaction times and standard deviation τ equal to the standard deviation of the eleven sample means $\bar{\bar{x}}_s$, s=1, …, 11.

Find the posterior distribution for θ_s, s=1, …, 11. Compare with your results in Problem 2.

Solution:

The posterior distribution for θ_s is normal, with:

- Mean $\mu_s^* = \frac{\mu/\tau^2+30\bar{x}_s/\sigma^2}{1/\tau^2+30/\sigma^2} = \frac{5.72/0.117^2+30\bar{x}_s/0.157^2}{1/0.117^2+30/0.157^2} = \frac{419.0+30\bar{x}_s/0.157^2}{1290.8}$
- Standard deviation $\tau^* = 1/\sqrt{1/\tau^2 + 30/\sigma^2} = 1/\sqrt{1290.8} = 0.0278$

The table below shows the eleven posterior means and 95% credible intervals.

Subject	Posterior Mean	2.5%	97.5%
θ_1	5.731	5.677	5.786
θ_2	5.886	5.832	5.941
θ_3	5.711	5.656	5.765
θ_4	5.706	5.651	5.760
θ_5	5.579	5.524	5.633
θ_6	5.794	5.740	5.849
θ_7	5.859	5.804	5.913
θ_8	5.588	5.534	5.643
θ_9	5.551	5.497	5.606
θ_{10}	5.774	5.720	5.829
θ_{11}	5.716	5.662	5.771

The values are very close to the values obtained by Gibbs sampling. If the goal is inference about the means, then the empirical Bayes analysis provides results that are very close to the full Bayesian analysis.

***Problem 4*:**

For this problem, you were asked to assume that the natural logarithms of the response times for each non-schizophrenic subject are independent and normally distributed with person-specific mean θ_j (j = 1,…11) and common variance σ^2. Discuss whether you think these assumptions are reasonable. Consider both theoretical arguments and analysis of the data.

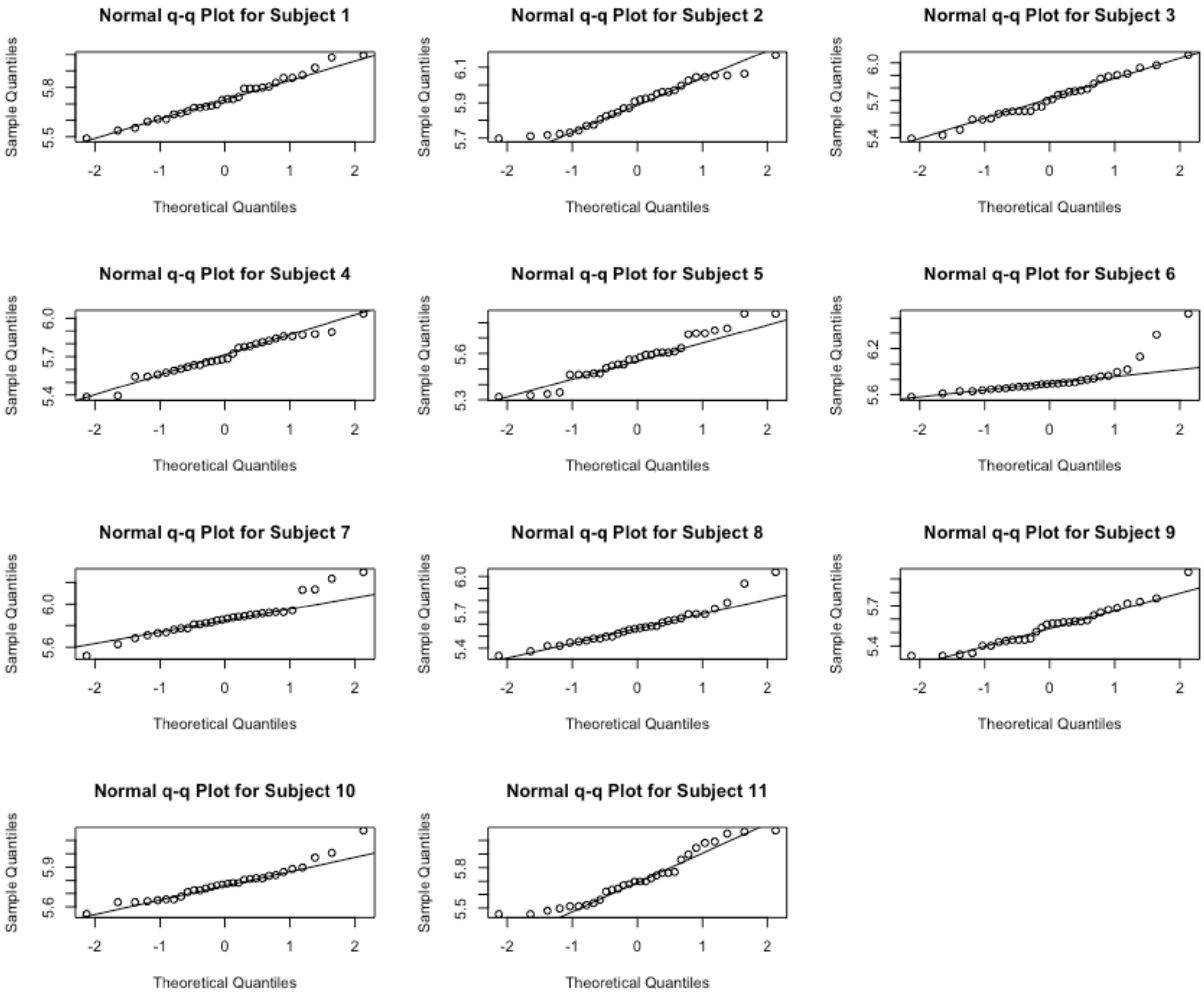
***Solution*:**

To investigate whether the observations are normally distributed, I did eleven normal quantile-quantile plots, as shown in the plots on the next page.

Several of the plots do not look normal; for example, subjects 6, 7 and 8. I also did a Shapiro-Wilks test for normality. The observations for Subjects 6, 7, 8 and 11 have p-values less than 0.05. Subject 6 has a p-value of about 10^{-6}; Subjects 7, 8 and 11 have p-values of 004, 0.03 and 0.04, respectively. So there is some evidence against normality for four subjects; the evidence is strong for Subject 11. If we are interested only in inferences about the mean and not about the tails of the distribution, assuming normality even for these subjects may not be too problematic.

A common error is to do a single q-q plot with all the observations. If the means are different, the collection of all observations will not be normal even if the individual plots are normal.

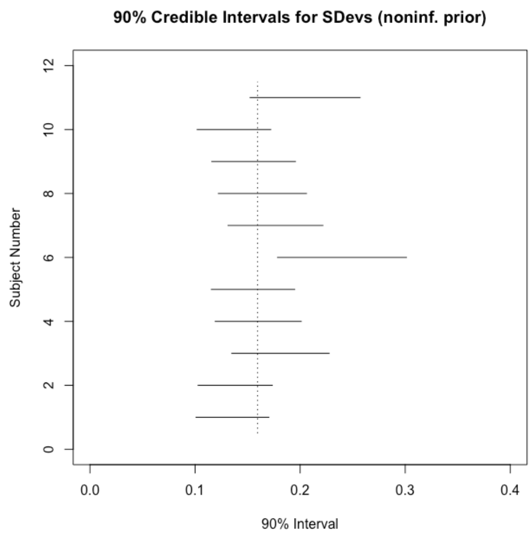
We have strong evidence that the means are not all the same. This is obvious from the caterpillar plot. Several of the credible intervals have no overlap with some of the other credible intervals.

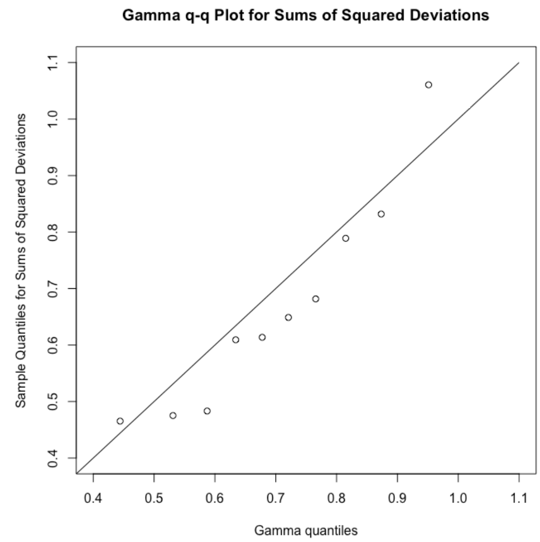


We can evaluate the assumption of equal standard deviations by using a hypothesis test such as Bartlett's test for equal variances.[1] Bartlett's test gives a p-value of 0.036, which is evidence against the hypothesis that all standard deviations are equal. However, this test could be affected by the non-normal data. Using the nonparametric Fligner-Kileen test, the p-value is 0.48, which does not reject the equal variance assumption.

Another way to investigate this question is to use a non-informative prior distribution to find posterior distributions for all the precisions, and find 95% credible intervals for the standard deviations. A plot of these intervals is shown below, with the average of the sample standard deviations shown as a vertical dotted line. We can see that the interval for Subject 6 does not overlap at all with the interval for subjects 1, 2 and 10.

[1] https://en.wikipedia.org/wiki/Bartlett%27s_test





Another diagnostic we can use is a q-q plot. If all the variances are equal, all the sums of squared deviations are gamma random variables with shape 14.5 and scale $2\sigma^2$, where σ^2 is the common variance. The right-hand plot shows a gamma q-q plot for the sample sums of squared deviations. The plot looks fairly linear. The largest sum of squared deviations corresponds to Subject 6, and is larger than expected, but not extremely so. From this plot, the equal variance assumption might be acceptable.

To summarize, all the standard deviations are all pretty similar, and if our goal is inference about the means, we might not go too far wrong by assuming they are equal. If we assume equal variances, then all credible intervals for the means are the same length. This might understate the uncertainty for subject 6, whose standard deviation may be larger than the others. On the other hand, the sample sum of squares for subject 6 doesn't seem exceptionally high when examining the q-q plot, and might just be a reflection of sampling variation, especially given the obvious non-normality of subject 6's data.