Name: ____

AMS 206B Final

Due: 5pm on March 21st (Wednesday), 2018

- Please typeset your solutions. Typesetting with latex is highly recommended. Partially handwritten solutions are acceptable, but handwritten only for mathematical equations.
- You need to turn in a **hard copy of your solutions.** If you type in your solutions and organize all well, you may email your file (not files) to me. If not, please drop yours in my office. You may turn in your solution before the deadline. If I am not in office, please slide your solution under the door. Please organize and present the material in the best possible way of answering all the questions. Be informative but concise and annotate all relevant figures and tables.
- **You must show all your work and justify all steps.**
- Email your code before the deadline. Add comments in your codes briefly.
- **Because this is an exam, you may not discuss the questions or solutions with any other students.** Do not share with anyone any information or comments about your findings or the models and methods you use.

The data includes quality control measurements from 6 machines in a factory; five measurements were obtained from each machine. You may load the dataset into R workspace using the following commands;

```
load("Machine1.RData")
> Machine
     machine measurements
 [1,]       1           83
 [2,]       1           92
 [3,]       1           92
 [4,]       1           46
 [5,]       1           67
 [6,]       2          117
 [7,]       2          109
```

If you are not in the same directory where "Machine1.RData" is saved, you may need to include the path to the directory,

```
load("/Path/Machine1.RData")
```

We let y_{ij} denote the j-th measurement from machine i, where $i = 1, \ldots, n$ with $n = 6$ and $j = 1, \ldots, m_i$ with $m_1 = \ldots = m_6 = 5$.

1. (10 pts) Perform explanatory analysis (e.g. making histograms, computing simple summary statistics for each machine) to examine if there are some differences in measurements between the machines. Summarize your findings.

2. (50 pts) Consider the following model (**Model 1**),

$$
\begin{aligned}
y_{ij} \mid \theta_i, \sigma^2 &\overset{indep}{\sim} \mathrm{N}(\theta_i, \sigma^2) \\
\theta_i \mid \mu, \tau^2 &\overset{iid}{\sim} \mathrm{N}(\mu, \tau^2).
\end{aligned}
$$

We also assume $\sigma^2 \mid \nu_0, s_0^2 \sim \mathrm{IG}(\nu_0/2, s_0^2/2)$, $\mu \sim \mathrm{N}(\mu_0, \omega^2)$ and $\tau^2 \sim \mathrm{IG}(a_\tau, b_\tau)$, where ν_0, s_0^2, μ_0, ω^2, a_τ and b_τ are fixed.

   (a) Write down the joint posterior distribution up to proportionality.

   (b) Provide the full conditionals for all parameters to run a Markov chain Monte Carlo (MCMC) simulation.

   (c) Specify values for the fixed hyperparameters and run MCMC simulations. Provide the values in a table and explain your choice. Report a summary of the posterior inference on parameters $\{\theta_i, \sigma^2, \mu, \tau^2\}$ such as point estimates and interval estimates. Provide relevant figures and/or tables.

   (d) Compare your posterior estimates of θ_i and σ^2 to their sample estimates (sample mean $\bar{y}_i = \sum_j y_{ij}/m_i$, machine specific sample variance $s_i^2 = \sum_j (y_{ij} - \bar{y}_i)^2/(m_i - 1)$ and overall variance $s^2 = \sum_{i,j}(y_{i,j} - \bar{y}_i)^2/(\sum_i m_i - 1)$). Comment.

3. (50 pts) Modify **Model 1** and consider the following model (**Model 2**),

$$\begin{aligned} y_{ij} \mid \theta_i, \sigma_i^2 &\overset{indep}{\sim} \mathrm{N}(\theta_i, \sigma_i^2) \\ \theta_i \mid \mu, \tau^2 &\overset{iid}{\sim} \mathrm{N}(\mu, \tau^2). \end{aligned}$$

We also assume $\sigma_i^2 \mid \nu_0, s_0^2 \overset{iid}{\sim} \mathrm{IG}(\nu_0/2, s_0^2/2)$, $s_0^2 \sim \mathrm{Gamma}(a_s, b_s)$, $\mu \sim \mathrm{N}(\mu_0, \omega^2)$ and $\tau^2 \sim \mathrm{IG}(a_\tau, b_\tau)$, where ν_0, a_s, b_s, μ_0, ω^2, a_τ and b_τ are fixed.

(a) Write down the joint posterior distribution up to proportionality.

(b) Provide the full conditionals for all parameters to run a MCMC simulation.

(c) Specify values for the fixed hyperparameters and run MCMC simulations. Provide the values in a table and explain your choice. Report a summary of the posterior inference on parameters $\{\theta_i, \sigma_i^2, \mu, \tau^2\}$ such as point estimates and interval estimates. Provide relevant figures and/or tables.

(d) Compare your posterior estimates of θ_i and σ_i^2 to their sample estimates (sample mean $\bar{y}_i$, sample variance s_i^2 and overall variance s^2). Comment.

4. (20 pts) Compare the inferences on parameters $\{\theta_i, \sigma_i^2 \text{ (or } \sigma^2), \mu, \tau^2\}$ under the two models. Also, perform a model comparison. Which model fits the data better? Explain why.

5. (20 pts) Report (i) the posterior distribution of the mean of the quality measurements of the sixth machine, (ii) the predictive distribution for another quality measurements of the sixth machine, and (iii) the posterior distribution of the mean of the quality measurements of the seventh machine. Compare the inferences under the two models.