Name: ____

STAT 206B Final
Due: 12:00pm on March 18th (Friday), 2022

- Typesetting with latex is highly recommended. Partially handwritten solutions are acceptable, but handwritten is allowed only for mathematical equations.
- Organize your solution well. Organize and present the material in the best possible way of answering the questions. Be informative but concise and annotate all relevant figures and tables. Poor organization may result in not getting full credits.
- Email your solution file (a single .pdf file) to me (juheelee@soe.ucsc.edu). You may turn in your solution before the deadline.
- Email your codes with your solution in a separate file. Your codes will be part of grading. Remove lines not needed for grading and add comments in your codes briefly. Your codes should be executable. Semi-automatic programming languages such as BUGS or Stan are not allowed.
- **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.
- Page limit: 10 pages for the report (the codes are not counted for this).

We use a weight gain dataset. For $n = 30$ women, weight gain during pregnancy is measured over time. y_{ij} represents weight gain in pounds at time in weeks t_{ij}, $j = 1, \ldots, m_i$ for subject i, $i = 1, \ldots, n$. The inferential goal of interest is to explain the relationship of weight gain with time and estimate a weight gain curve of a pregnant woman over time. The data file `pregnency-data.csv` is separately attached. Below illustrate first 10 observations, where the first seven observations are weight gain measurements from subject 1 at time points 12.08, 16.46, 20.76, 24.54, 28.91, 32.70, 36.97 (so $m_1 = 7$);

```
> dat_1[1:10,]
    X patient visit     time   weight
1   1       1     1 12.07580 10.12724
2   2       1     2 16.45865 14.80822
3   3       1     3 20.76493 17.49366
4   4       1     4 24.53566 19.74380
5   5       1     5 28.90688 23.36855
6   6       1     6 32.69701 24.00172
7   7       1     7 36.96997 24.85463
8   8       2     1 12.21765 13.68243
9   9       2     2 16.88070 18.38908
10 10       2     3 20.33713 22.79842
>
.....
```

1. (15 pts) Using quantitative and graphical tools summarize the main features of this dataset.

2. Consider the following model,

$$y_{ij} \mid \beta_{0i}, \beta_1, \beta_2, \sigma^2 \overset{indep}{\sim} \mathrm{N}\left(\frac{\beta_{0i}}{1 + \beta_1 \exp(\beta_2 t_{ij})}, \sigma^2\right), \quad i = 1, \ldots, n \text{ and } j = 1, \ldots, m_i. \tag{1}$$

   We assume the following distributions as prior distributions;

$$\begin{aligned} \beta_{0i} \mid \bar{\beta}_0, u_0^2 &\overset{iid}{\sim} \mathrm{N}(\bar{\beta}_0, \tau^2), \quad i = 1, \ldots, n, \\ \beta_p \mid \bar{\beta}_p, u_p^2 &\overset{iid}{\sim} \mathrm{N}(\bar{\beta}_p, u_p^2), \quad p = 1, \text{ and } 2, \\ \sigma^2 \mid a_\sigma, b_\sigma &\sim \mathrm{IG}(a_\sigma, b_\sigma), \end{aligned}$$

   where $\bar{\beta}_p$, u_p^2, $p = 1, 2$, a_σ and b_σ are fixed. We further let

$$\begin{aligned} \bar{\beta}_0 \mid \mu_0, v^2 &\sim \mathrm{N}(\mu, v^2), \\ \tau^2 \mid a_\tau, b_\tau &\sim \mathrm{IG}(a_\tau, b_\tau). \end{aligned}$$

   Here, μ, v^2, a_τ and b_τ are fixed. We may read the wikipedia page on a logistic curve (https://en.wikipedia.org/wiki/Logistic_function) for more understanding of the assumed mean function.

   (a) (10 pts) Write down the joint posterior distribution of all random model parameters up to a proportionality constant.

   (b) (15 pts) Describe a Markov chain Monte Carlo (MCMC) algorithm in details for posterior computation for the model. Provide the full conditionals needed for your algorithm.

(c) (10 pts) Specify values of the fixed hyperparameters and justify your choices.

(d) (20 pts) Implement your algorithm in (b) and run MCMC simulations. Assess mixing and convergence of your Markov chain. Provide evidence that there is no poor mixing or approximate convergence is reached.

Note: You may try various values of the fixed hyperparameters for a sensitivity analysis. Select your best set of the hyperparameter values for your report.

(e) (10 pts) Report a summary of the posterior inference on all random parameters such as their point estimates and interval estimates. Provide relevant figures and/or tables. Provide a brief interpretation of your results in layman's terms and discuss.

(f) (15 pts) Perform the model checking as described below;

i. Compute a posterior predictive distribution of weight gain for subject 1 (i.e., $i = 1$) at each of the time points, t_{1j}, $j = 1, \dots, m_i$ with $m_1 = 7$.

ii. Compare the posterior predictive distributions to the actual observed values of the woman, and evaluate the fit of the model to the data. Provide relevant figures and/or tables to support you findings.