Stat 532   **Name:** _____

Take home final

Due 12/04/17 at 9:00 AM

_____

For the take home exam, you may use the textbook, any course materials provided on D2L, home-works, and labs. You **may not** discuss questions or work together with classmates. You are welcome to contact the instructor with any questions related to better explanation or understanding of the questions themselves. For complete (and partial credit) please show all work, whether that be by hand or printed R code.

1. Suppose you have been hired by CapitalBikeshare, a bike share company based in the Washington D.C. area, to construct a model of bike rentals. You have been provided a dataset http://www.math.montana.edu/ahoegh/teaching/stat532/data/biketrips2017Final.csv that contains the number of daily bike rentals from seven stations.

   (a) (4 points) As a first model, write out a hierarchical normal model to estimate the mean number of bike rentals at each station and specify all necessary prior distributions.

   (b) (4 points) Use an MCMC technique to fit this model and create a table or figure to display the credible intervals for the mean number of bikes rented at each station.

   (c) (4 points) Now assume you are asked to describe your work to a set of executives from CapitalBikeshare. Script (write out) your presentation of the model and results you can include the equivalent of one slide for graphics or a table to accompany your speech.

   (d) (4 points) After hearing your description, one executive asks *'Are your results different from a set of independent models that take average number of rentals at each station? and if so, why is your approach better?'*. Address this question.

   (e) (4 points) Assume another executive at CapitalBikeshare asks why you didn't use a p-value in your analysis. Describe how you would answer that question.

2. Now consider a dataset containing daily counts at the Lincoln Memorial.
   http://www.math.montana.edu/ahoegh/teaching/stat532/data/biketrips2017Lincoln.csv
   Fit a Poisson regression model (generalized linear model) using day of the week as a covariate.

   (a) (4 points) Write out the model and necessary prior distributions for this model.

   (b) (4 points) Detail *and* implement an MCMC procedure to fit this model.

   (c) (4 points) Summarize the posterior distribution with a table or visually with a plot and describe your results.

   (d) (4 points) Create a figure for the predictive distribution for the mean count of bikes on Wednesdays.

   (e) (4 points) Implement *and* detail a procedure to verify that your prior and sampling model are reasonable in this case.

3. Regardless of the result for the previous question, now assume you decide to fit a negative binomial generalized linear model using day of week as a covariate to model the count of rentals at the Lincoln Memorial.

   (a) (4 points) Write out the model and include necessary priors.

   (b) (4 points) Now sketch out how an MCMC approach could be implemented. For each parameter write out the full conditional distribution up to proportionality and describe how you would sample that parameter. Note you do not have to implement this approach.

4. (6 points) Describe and justify your personal philosophy statistical philosophy. How has this changed (or not) since the start of the semester. (roughly half a page)

5. (6 points) The ASA statement on p-values highlights 6 principles about p-values
   http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108:

   1. P-values can indicate how incompatible the data are with a specified statistical model.

   2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

   3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

   4. Proper inference requires full reporting and transparency.

   5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

   6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

   Identify two of these principles and discuss how they are different from a Bayesian perspective. (roughly half a page)