STAT 491: Final Exam
Due: May ?
Name:

Please turn in the exam to D2L and include the R Markdown code, SAS code *and either* a Word or PDF file with output. While the exam is open book, meaning you are free to use any resources from class, this is strictly an individual endeavor and **you should not discuss the problems with anyone outside the course instructor including class members.** The instructor will answer questions related to the data, expectations, and understanding of the exam, but will not fix or troubleshoot broken code.

(60 points Indeed.com Data Analysis)

This exam will focus on small dataset containing information from Indeed.com, which can be accessed using http://www.math.montana.edu/ahoegh/teaching/stat491/data/bzn_jobs.csv.

```
bzn.jobs <- read.csv('http://www.math.montana.edu/ahoegh/teaching/stat491/data/bzn_jobs.csv')
head(bzn.jobs)
```

```
##   jobAgeDays              normTitle estimatedSalary localClicks
## 1          4 retail sales associate           25200           7
## 2          3           truck driver           39700           0
## 3          2 retail sales associate           19800           3
## 4          0       registered nurse           59600           9
## 5          0       registered nurse           72000           0
## 6          5 retail sales associate           22100          17
```

This dataset contains the following variables:

```
- jobAgeDays: number of days the job has been posted on Indeed.com
- normTitle: name of job position
- estimatedSalary: estimated annual salary
- localClicks: number of people clicking on job posting
```

Question 1. (36 points) - Bayesian Linear Models

For this question we will fit a regression analysis to model estimated Salary. For full credit you need to consider all other variables as predictors.

a. (4 points)

Explain the purpose of this model - you can assume you talking to a freshman in high school.

b. (4 points)

Select and Justify a Sampling Model

c. (4 points)

Write out the Linear Combination of Predictors for your model and justify this selection

d. (4 points)

State and Justify Priors Used for this Model

e. (4 points)

Use JAGS to fit the Posterior Distribution for this Model and Summarize the Results

f. (4 points)

Use your model to construct a posterior predictive distribution for the estimatedSalary of a new job with: jobAgeDays = 1, normTitle = registered nurse, and localClicks = 11.

g. (4 points)

Explain the results of this model - you can assume you talking to a freshman in high school.

h. (4 points)

Discuss the differences between using normTitle as a nominal value and fitting a hierarchical regression model with normTitle as the group. You don't need to fit these models, but be specific in your discussion.

i. (4 points)

Discuss the differences between using a t-distribution and a normal distribution on the sampling model. How would you decide which was more appropriate?

Question 2. (24 points) - Bayesian Poisson Regression

Now assume the goal is to model the number of localClicks for each job using Poisson regression.

a. (4 points)

Explain the purpose of this model - you can assume you talking to an executive considering whether to post a job on Indeed.com. (Note: companies posting featured jobs pay Indeed based on the number of clicks on the job posting)

b. (4 points)

Use a Poisson distribution as the sampling model and write out the linear combination of predictors for your model and justify this selection

c. (4 points)

State and justify priors used for this model

d. (4 points)

Use JAGS to fit the Posterior Distribution for this Model and Summarize the Results

e. (4 points)

Use your model to construct a posterior predictive distribution for the localClicks of a new job with: jobAgeDays = 1, normTitle = registered nurse, and estimatedSalary = 70000.

f. (4 points)

Explain the results of this model - you can assume you talking to an executive considering whether to post a job on Indeed.com.