Activity 7

Name here

This analysis 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')
```

```
## Rows: 30 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): normTitle
## dbl (3): jobAgeDays, estimatedSalary, localClicks
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

This dataset contains the following variables:

```
- jobAgeDays: number of days the job has been posted on Indeed.com
- normTitle: name of job position (registered nurse, sales associate, truck driver)
- estimatedSalary: estimated annual salary
- localClicks: number of people clicking on job posting
```

Bayesian ANOVA

For this question we will fit a regression analysis (ANOVA) to model estimated salary across three different job types.

a. Data Viz Create a figure of salary by `normTitle`. It is good practice to show all data points.

b. Interpret the following R output.

```
anova_fit <- lm(estimatedSalary ~ normTitle - 1, data = bzn_jobs)
summary(anova_fit)
```

```
##
## Call:
## lm(formula = estimatedSalary ~ normTitle - 1, data = bzn_jobs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12362.5  -3008.7    181.2   2626.9  12325.0
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## normTitleregistered nurse         61575       1706   36.10  < 2e-16 ***
```

```
## normTitleretail sales associate    22310       1869   11.94 2.78e-12 ***
## normTitletruck driver              38862       2089   18.60  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5909 on 27 degrees of freedom
## Multiple R-squared:  0.9852, Adjusted R-squared:  0.9835
## F-statistic: 597.2 on 3 and 27 DF,  p-value: < 2.2e-16
```

```
confint(anova_fit)
```

```
##                                    2.5 %   97.5 %
## normTitleregistered nurse       58075.08 65074.92
## normTitleretail sales associate 18476.03 26143.97
## normTitletruck driver           34575.99 43149.01
```

c. Select and Justify a sampling model for your response.

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

e. State and Justify Priors Used for your Model

e. Modify existing JAGS code to fit this model.

```
indicator_data <- model.matrix(estimatedSalary~normTitle - 1, data = bzn_jobs)
```

```
model_anova<- "model{
  # Likelihood
  for (i in 1:n){
    y[i] ~ dnorm(beta1 * x1[i] + beta2 * x2[i] + beta3 * x3[i], 1/sigma^2)
  }

}"
```

f. (4 points) Using your JAGS code to fit the Posterior Distribution for this Model and print the results

g. Visualize your results in some fashion.

h. Compare your interval results, in part f, with those from part b

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