Paper 1 Assignment
Bernoulli Response

In 2003, researchers published the results of a study to compare sex differences related to the incidence of concussion in college athletes in three seasons: 1997-1998, 1998-1999, and 1999-2000. In this paper, you will explore logistic regression models to assess the relationship between concussion and sport, adjusting appropriately for potential confounding variables. See the class syllabus and homework rubric for grading details.

The data can be read in as follows; the data set is also available for download on the Moodle page, or can be accessed through the class Github page here: https://github.com/marievozanne/marievozanne.github.io/blob/master/stat340_f2020/data/concussion.csv

```r
library(tidyr)

## Data source: http://users.stat.ufl.edu/~winner/data/concussion.txt
concussion <- read.table("http://users.stat.ufl.edu/~winner/data/concussion.dat")
names(concussion) <- c("Sex", "Sport", "Year", "Concussion", "Count")
head(concussion)
```

```
##        Sex  Sport Year Concussion Count
## 1 Female Soccer 1997          0 24930
## 2 Female Soccer 1997          1    51
## 3 Female Soccer 1998          0 22887
## 4 Female Soccer 1998          1    47
## 5 Female Soccer 1999          0 27107
## 6 Female Soccer 1999          1    60
```

```r
## Require 0/1 as response for glm
concussion <- concussion %>%
  uncount(Count)
```

References

T. Covassin, C.B. Swanik, M.L. Sachs (2003). Sex Differences and the Incidence of Concussions Among Collegiate Athletes, *Journal of Athletic Training*, Vol. (38)3, pp. 238-244

Data sets - Statistics, University of Florida. http://users.stat.ufl.edu/~winner/datasets.html