# Homework 3

Trevor Caughlin

January 29, 2018

## Question 1: What is the effect of cut quality on diamond price?

These data are the prices of 50,000 round cut diamonds (source: http://ggplot2.tidyverse.org/reference/diamonds.html )

Data includes: **Price** in US dollars ($326–$18,823), **Carat** (weight) of the diamond (0.2–5.01) , and **Cut** (quality of the cut; Fair, Good, Very Good, Premium, Ideal)

```
diamonds<-read.csv("diamond.csv")
head(diamonds)
```

```
##   price       cut carat
## 1   326     Ideal  0.23
## 2   326   Premium  0.21
## 3   327      Good  0.23
## 4   334   Premium  0.29
## 5   335      Good  0.31
## 6   336 Very Good  0.24
```

What is the effect of each cut on the price of a typical diamond (in dollars)?

As output, produce a boxplot of original data, parameter estimates, and 95% CI

## Question 2: Does education have an impact on contraception use?

These data represent contraception use in a sample of 1607 married women from Fiji.

Source: "Reference: Little, R. J. A. (1978). Generalized Linear Models for Cross-Classified Data from the WFS. World Fertility Survey Technical Bulletins, Number 5."

Data include: **age** (cohort of women), **education** (education level), **notUsing** (count of women in each sample not using contraception), **using** (count of women in each sample using contraception), and **total number of women in sample** (total count)

```
cuse<-read.csv("contraception.csv")
head(cuse)
```

```
##     age education notUsing using Total
## 1   <25       low       53     6    59
## 2   <25       low       10     4    14
## 3   <25      high      212    52   264
```

```
## 4   <25      high       50    10    60
## 5 25-29       low       60    14    74
## 6 25-29       low       19    10    29
```

Use a binomial glm to test the hypothesis that increased education promotes contraception use.

As output, produce a boxplot of data, parameter estimates, and 95% CI (and a short sentence on whether you reject the null hypothesis or not).

## Question 3

A controversial recent paper claimed that female-named hurricanes led to more deaths than male-named hurricanes (also known as himmicanes), because people do not take hurricanes as seriously as himmicanes. You can find a link to the paper here: http://www.pnas.org/content/111/24/8782.short

Citation: Jung, Kiju, Sharon Shavitt, Madhu Viswanathan, and Joseph M. Hilbe. 2014. "Female Hurricanes Are Deadlier than Male Hurricanes." Proceedings of the National Academy of Sciences 111 (24): 8782-87. doi:10.1073/pnas.1402786111.

Here are the data used for their paper:

```
hurricanes<-read.csv("Hurricane Dataset.csv")
head(hurricanes)
```

```
##   Year     Name  MasFem MinPressure_before Minpressure_Updated.2014
## 1 1950     Easy 6.77778                958                       960
## 2 1950     King 1.38889                955                       955
## 3 1952     Able 3.83333                985                       985
## 4 1953  Barbara 9.83333                987                       987
## 5 1953 Florence 8.33333                985                       985
## 6 1954    Carol 8.11111                960                       960
##   Gender_MF Category alldeaths  NDAM Elapsed.Yrs Source  ZMasFem
## 1         F        3         2  1590          63    MWR -0.00094
## 2         M        3         4  5350          63    MWR -1.67076
## 3         M        1         3   150          61    MWR -0.91331
## 4         F        1         1    58          60    MWR  0.94587
## 5         F        1         0    15          60    MWR  0.48108
## 6         F        3        60 19321          59    MWR  0.41222
##   ZMinPressure_A    ZNDAM
## 1       -0.35636 -0.43913
## 2       -0.51125 -0.14843
## 3        1.03765 -0.55047
## 4        1.14091 -0.55758
## 5        1.03765 -0.56090
## 6       -0.25310  0.93174
```

Re-analyze their data using Poisson glm, with a focus on **alldeaths** as a response variable and **Gender_MF** as a predictor variable. How do your results differ from Jung et al.'s conclusions? What could Jung et al. have done differently to inspire more confidence in their analyses?

## Question 4

Find one dataset from your own research that could be modelled as normal, binomial, or poisson regression (note: http://datadryad.org/ is a great source for existing datasets if you don't have your own data yet). Run a glm on your data and interpret results in terms of effect size, visualization, and statistical significance.