Homework 3 Juliette Rubin

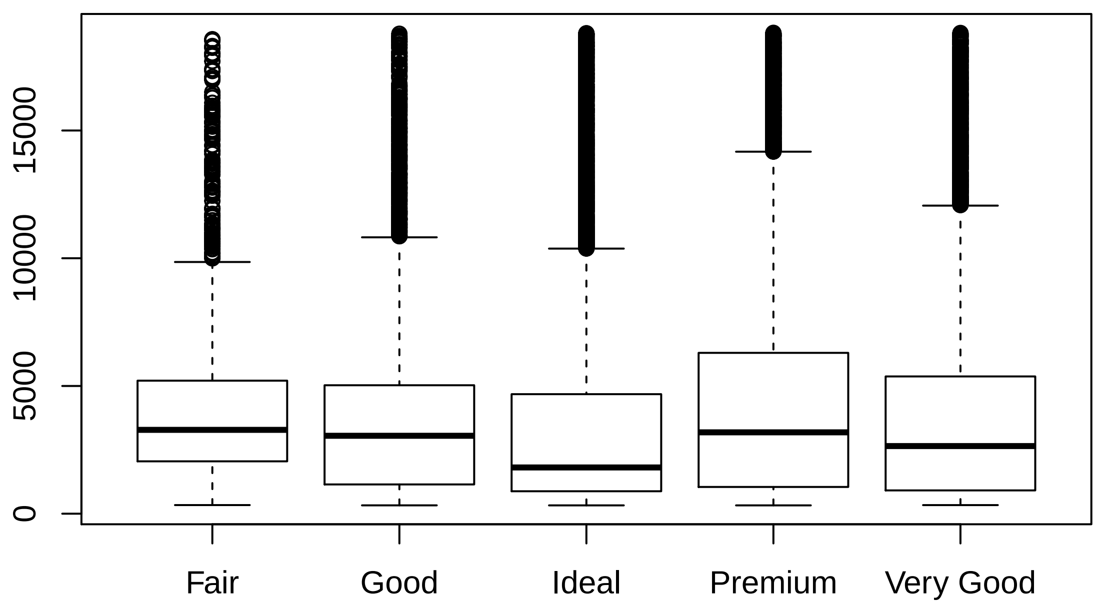
Juliette J. Rubin

February 14, 2019

Question 1

The effect of diamond cut on overall price. Because the response variable always has to be positive (can't have a diamond that costs negative dollars! Becuase capitalism!)and there is an upper bound (most costly diamond is $18,823), additionally, the variance is not equal to the mean of these values, so it should not be run as a Poisson. When we visualize the data using boxplots, it seems that the data might be overdispersed (there is more variance than expected, as evidenced by all the outliers and the general right skew of the data. We will therefore use a negative binomial.

```
diamond<-read.csv("diamond.csv")
#visualizing data
boxplot(price~cut,data=diamond)
```




```
diamond_mod<-glm.nb(price~cut, data=diamond)

#see below for interpretation
coef(diamond_mod)
```

```
##  (Intercept)      cutGood     cutIdeal   cutPremium cutVery Good
##    8.3799424   -0.1038367   -0.2316292    0.0504411   -0.0904632
```

```r
#see below for interpretation
confint(diamond_mod)
```

```
## Waiting for profiling to be done...

##                     2.5 %      97.5 %
## (Intercept)   8.335159513  8.42540460
## cutGood      -0.156067341 -0.05206239
## cutIdeal     -0.278721273 -0.18516569
## cutPremium    0.002457342  0.09782480
## cutVery Good -0.138792710 -0.04272251
```

```r
#There is a 9.85% relative decrease between the cost of a fair and a good diamond
exp(-0.1038)
```

```
## [1] 0.9014056
```

```r
1-0.9015
```

```
## [1] 0.0985
```

```r
#There is a proportional decrease of costy by 20.7% between the fair and ideal cuts
exp(-0.232)
```

```
## [1] 0.7929461
```

```r
1-0.793
```

```
## [1] 0.207
```

```r
#There is a 5.2% increase in the number of dollars that the premium cut costs in relation to the fair c
exp(0.0504)
```

```
## [1] 1.051692
```

```r
1-1.0517
```

```
## [1] -0.0517
```

```r
#There is a 60% decrease in relative cost between the very good cut and the fair cut
exp(-0.9046)
```

```
## [1] 0.4047037
```

```r
1-0.4005
```

```
## [1] 0.5995
```

These are the intercepts that we find for each category of diamond on price (Intercept) cutGood cutIdeal cutPremium cutVery Good 8.3799424 -0.1038367 -0.2316292 0.0504411 -0.0904632

This seems to indicate that the initial cost of getting into diamon buying is so high that you then are paying relatively less to get an excellently cut diamond than simply buying the entry level diamond.

And here are the confidence intervals. As we can see, the effect of cut on price is significant in each case 2.5 % 97.5 % (Intercept) 8.335159513 8.42540460 cutGood -0.156067341 -0.05206239 cutIdeal -0.278721273 -0.18516569 cutPremium 0.002457342 0.09782480 cutVery Good -0.138792710 -0.04272251

Question 2

We are addressing the question of education level affects contraception use by women in Fuji. As you will not below, we find a surprising result that education level does not have a significant relationship with contraception use :

```
contra<-read.csv("contraception.csv")

#finding the proporation of people using contraception
prop_contra<-contra$using/contra$Total

#making a data matrix of people using and not
contra_mat<-cbind(contra$using,contra$Total-contra$using)

#making a binomial model
contraModel<-glm(contra_mat~contra$education, family="binomial")

confint(contraModel)
```

```
## Waiting for profiling to be done...
```

```
##                          2.5 %     97.5 %
## (Intercept)         -0.9460962 -0.6766394
## contra$educationlow -0.1239481  0.3078275
```

```
#the slope crosses zero therefore there is not a significant relationship
#the 95% conf int is between
coef(contraModel)
```

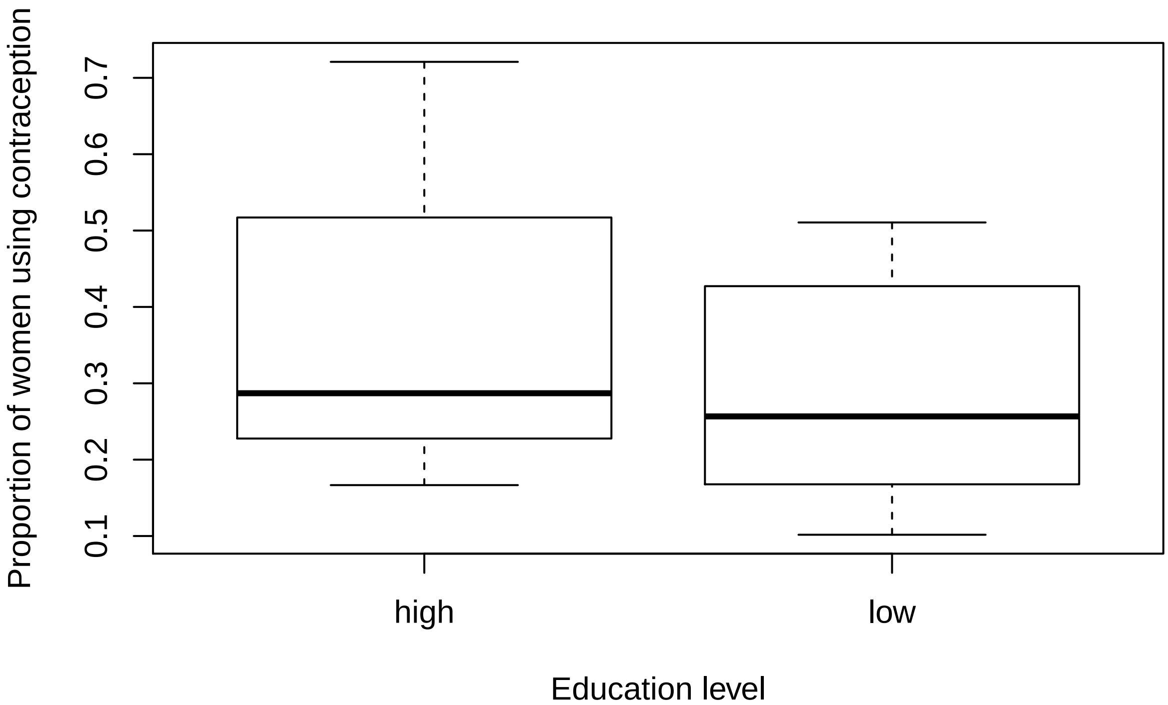
```
##         (Intercept) contra$educationlow
##         -0.81020374          0.09248529
```

```
#intercept: -0.81, 0.092
plogis(-0.81)
```

```
## [1] 0.3078905
```

```
#at baseline (i.e., with no education, 31% of women are using contraception)

boxplot(prop_contra~contra$education, xlab="Education level",ylab="Proportion of women using contracepti
```



Question 3

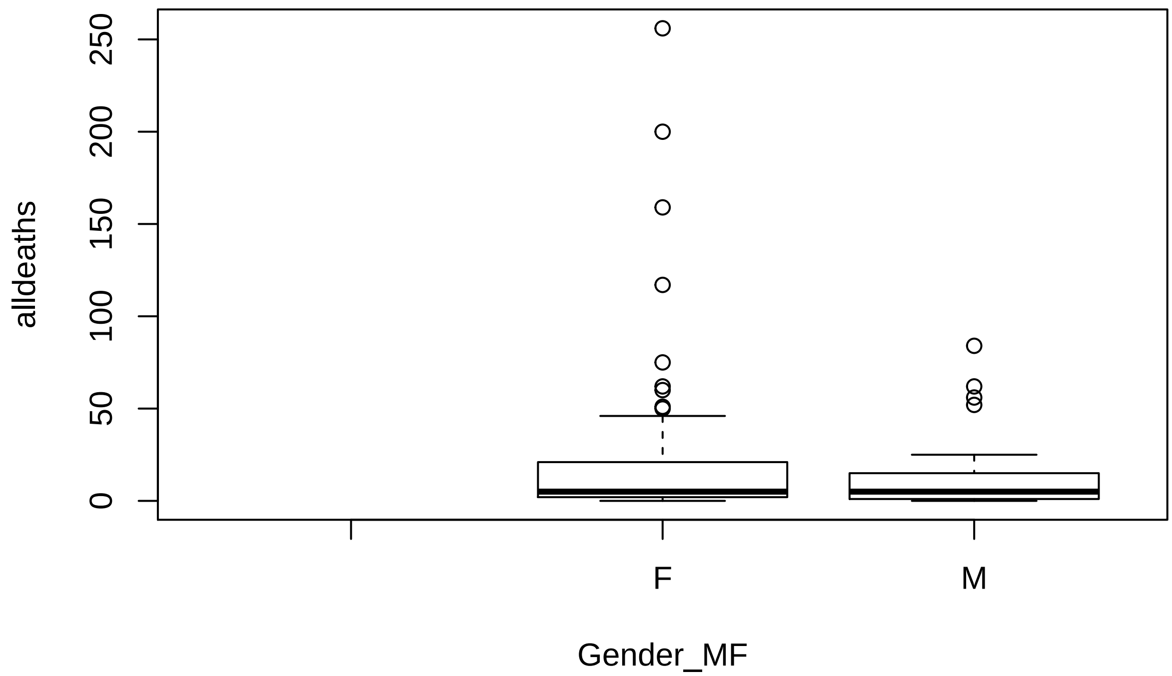
You can also embed plots, for example:

```
## Waiting for profiling to be done...

##                  2.5 %     97.5 %
## (Intercept)  3.1164152  3.2185581
## Gender_MFM  -0.6211542 -0.4056501

## (Intercept)  Gender_MFM
##   3.1679220  -0.5123354

## [1] -15.91828
```



While I ended up with the same result as this team: there does seem to be a significant relationship between hurricane name gender and number of deaths, I would be more confident in their analyses if they kept them a little more streamlined. They seemed to use a variety of different models, including negative binomial, poisson, poisson inverse Gaussian, etc. This makes it seem a little like a fishing expedition, rather than a serious inquiry.

Question 4

The following is a binomia regression of a preliminary data set where we pit bats against tiger moths (cycnia tenera) and either allowed the moths to make their anti-bat sound, or removed their ability to do so (ablating the sound producing structures; tymbals).

```
moth<-read.csv("cycnia_data.csv")

avoidance<-cbind(moth$avoid,moth$total-moth$avoid)

#making a binomial model
avoidance_model<-glm(avoidance~moth$Sound, family="binomial")

coef(avoidance_model)
```

```
## (Intercept)  moth$Sound
##   0.2384110   0.6088868
```

```
#intercept:0.238, slope:0.609

confint(avoidance_model)
```

```
## Waiting for profiling to be done...

##                  2.5 %    97.5 %
## (Intercept) -0.2731440 0.7604967
## moth$Sound  -0.1761827 1.4182678
```

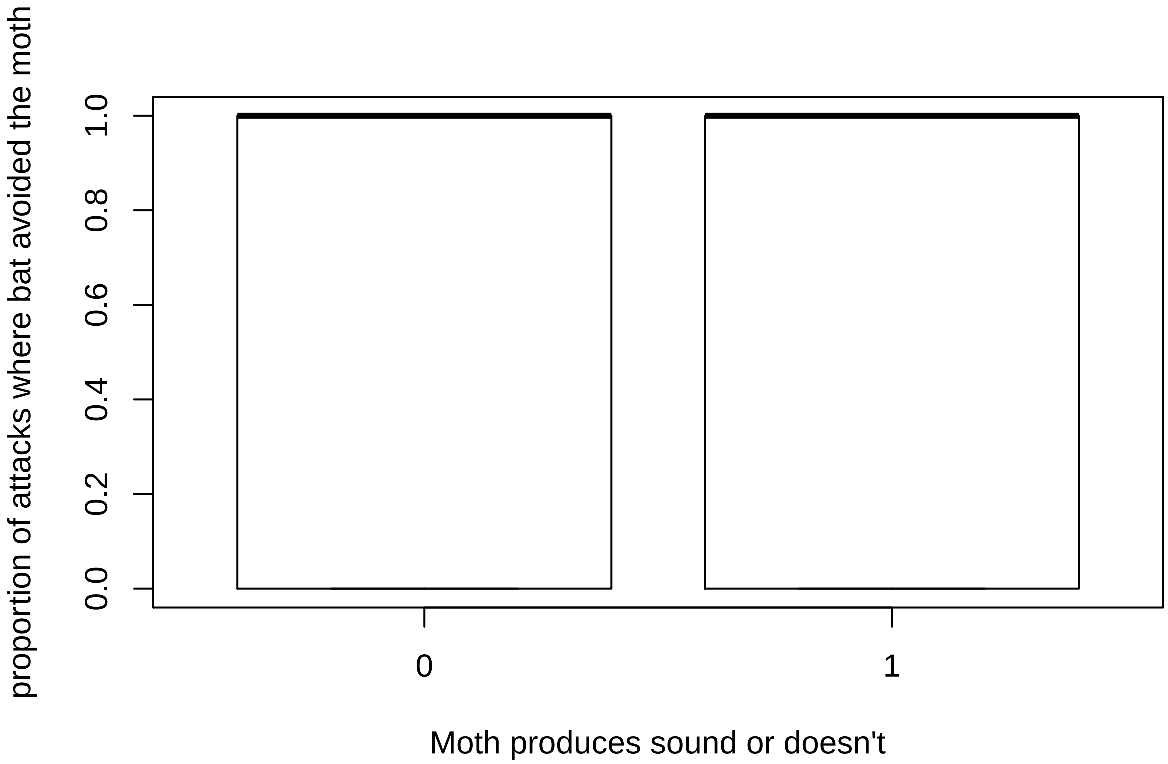
```
#The result is not significant. The 95% confidence interval overlaps zero.

plogis(0.238)
```

```
## [1] 0.5592207
```

```
#The baseline avoidance of a moth by a bat when the moth cannot make sound is 55.9%

prop_avoid<-moth$avoid/moth$total
boxplot(prop_avoid~moth$Sound, ylab="proportion of attacks where bat avoided the moth", xlab="Moth produ
```




```
#what am I doing wrong? Is it because they are both categorical? Is a Chi Square my only option?
```

We did not find an effect of sound production on bat avoidance of moths, based on this preliminary dataset