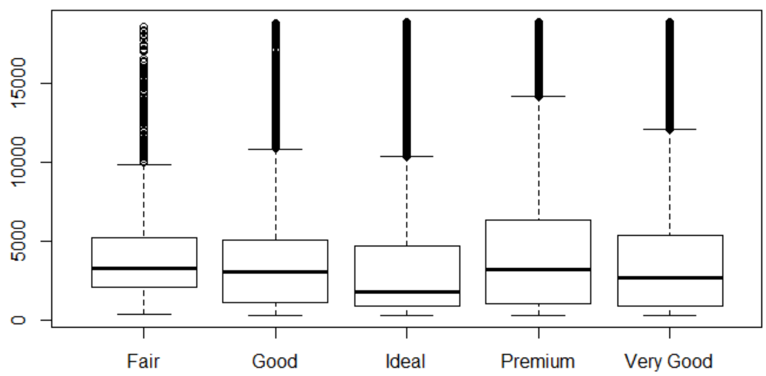
Megan Kelly-Slatten
2-14-19

Homework 3

Question 1: Diamond

```
> gem<-read.csv("R/diamond.csv")
> str(gem)
'data.frame':   53940 obs. of  3 variables:
 $ price: int  326 326 327 334 335 336 336 337 337 338 ...
 $ cut  : Factor w/ 5 levels "Fair","Good",..: 3 4 2 4 2 5 5 5 1 5 ...
 $ carat: num  0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...
> gemmod<-glm(gem$price~gem$cut, family="poisson")
> coef(gemmod)
     (Intercept) gem$cutGood gem$cutIdeal gem$cutPremium gem$cutVery Good
       8.3799424  -0.1038367  -0.2316292      0.0504411      -0.0904632
> confint(gemmod)
                     2.5 %      97.5 %
(Intercept)       8.37920242  8.38068216
gem$cutGood      -0.10470072 -0.10297248
gem$cutIdeal     -0.23240302 -0.23085517
gem$cutPremium    0.04966133  0.05122103
gem$cutVery Good -0.09125511 -0.08967112
> boxplot(gem$price~gem$cut)
```

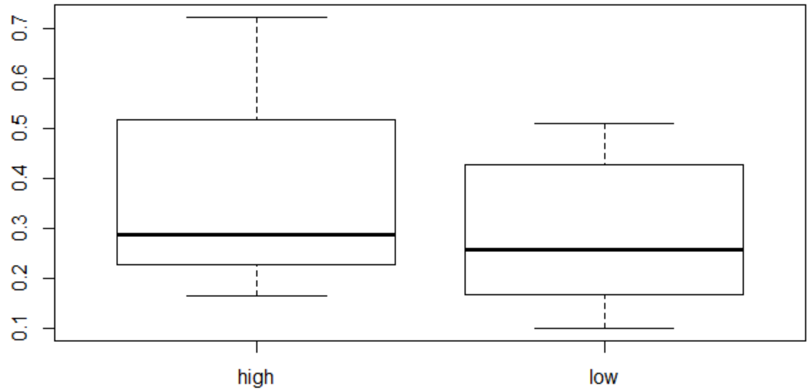



```
> exp(8.3799) ##baseline fair diamond is $4358
[1] 4358.573
> exp(8.3799-0.1038)-exp(8.3799) #good
[1] -429.7311
> exp(8.3799-0.2316)-exp(8.3799) #ideal
[1] -901.0767
> exp(8.3799+0.0504)-exp(8.3799) #premium
[1] 225.302
> exp(8.3799-0.0904)-exp(8.3799) #very good
[1] -376.7303
```

With a fair diamond cut the baseline price is $4358. When you move from a fair to good cut on average you decrease by $429. When you move from a fair cut to a very good cut on average you decrease by $376. When you move from a fair cut to an ideal cut on average you decrease by $901. When you move from a fair cut to a premium cut on average you increase by $225.

Question 2: Married, Fiji Woman

```
> wed<-read.csv("R/contraception.csv")
> str(wed)
'data.frame':	16 obs. of  5 variables:
 $ age      : Factor w/ 4 levels "<25","25-29",..: 1 1 1 1 2 2 2 2 3 3 ...
 $ education: Factor w/ 2 levels "high","low": 2 2 1 1 2 2 1 1 2 2 ...
 $ notUsing : int  53 10 212 50 60 19 155 65 112 77 ...
 $ using    : int  6 4 52 10 14 10 54 27 33 80 ...
 $ Total    : int  59 14 264 60 74 29 209 92 145 157 ...
> sex=wed$using/(wed$notUsing+wed$using)
> boxplot(sex~wed$education)
```



Level of education compared to proportion of women using contraception.

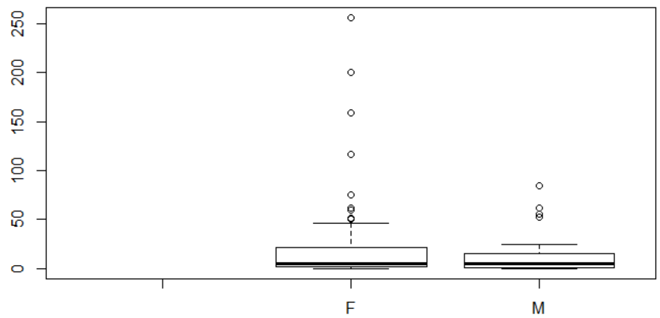
```
> cont=cbind(wed$using, wed$notUsing)
> femmod<-glm(cont~wed$education, family="binomial")
> coef(femmod)
     (Intercept) wed$educationlow 
     -0.81020374       0.09248529 
> plogis(-0.8102+0.0924)-plogis(-0.8102)
[1] 0.02002974
> confint(femmod)
                      2.5 %     97.5 %
(Intercept)      -0.9460962 -0.6766394
wed$educationlow -0.1239481  0.3078275
```

When woman are highly educated there is a 2% increase in women using contraception. However, this data does not appear to be significant because the CI95% overlap zero.

Question 3: STORMS

```
> storm<-read.csv("R/Hurricane Dataset.csv")
> str(storm)
'data.frame':	98 obs. of  14 variables:
 $ Year                     : Factor w/ 55 levels ""," Reanalysis with these d
ata did not alter the conclusions. The modifications were: Hurricane Easy (19
50:  958 t"| __truncated__,..: 4 4 5 6 6 7 7 7 8 8 ...
 $ Name                     : Factor w/ 84 levels "","Able","Agnes",..: 39 78
2 10 48 21 41 61 28 34 ...
 $ MasFem                   : num  6.78 1.39 3.83 9.83 8.33 ...
 $ MinPressure_before       : int  958 955 985 987 985 960 954 938 962 987 ...
 $ Minpressure_Updated.2014: int  960 955 985 987 985 960 954 938 962 987 ...
 $ Gender_MF                : Factor w/ 3 levels "","F","M": 2 3 3 2 2 2 2 2 2
2 ...
 $ Category                 : int  3 3 1 1 1 3 3 4 3 1 ...
```

```
 $ alldeaths                : int  2 4 3 1 0 60 20 20 0 200 ...
 $ NDAM                     : int  1590 5350 150 58 15 19321 3230 24260 2030 1
4730 ...
 $ Elapsed.Yrs              : int  63 63 61 60 60 59 59 59 58 58 ...
 $ Source                   : Factor w/ 4 levels "","http://www.nhc.noaa.gov/p
df/NWS-TPC-5.pdf",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ ZMasFem                  : num  -0.00094 -1.67076 -0.91331 0.94587 0.48108
...
 $ ZMinPressure_A           : num  -0.356 -0.511 1.038 1.141 1.038 ...
 $ ZNDAM                    : num  -0.439 -0.148 -0.55 -0.558 -0.561 ...
> boxplot(storm$alldeaths~storm$Gender_MF)
```

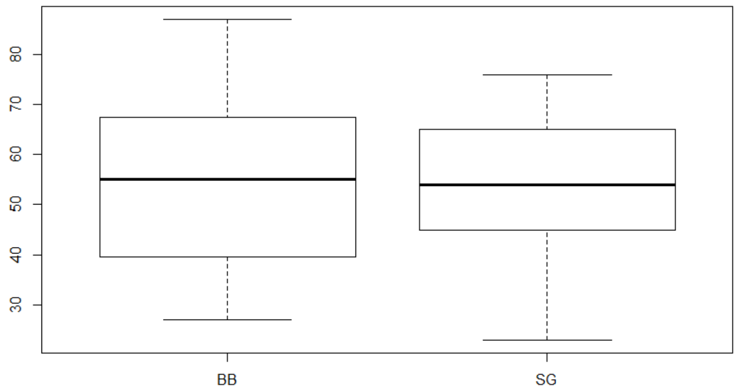



```
> stormmod<-glm(storm$alldeaths~storm$Gender_MF, family="poisson")
> coef(stormmod)
     (Intercept) storm$Gender_MFM
       3.1679220       -0.5123354
> confint(stormmod)
                      2.5 %     97.5 %
(Intercept)       3.1164152  3.2185581
storm$Gender_MFM -0.6211542 -0.4056501
> exp(3.1679-0.51233)-exp(3.1679)
[1] -9.524444
> library(MASS)
> storm2<-glm.nb(storm$alldeaths~storm$Gender_MF)
> coef(storm2)
     (Intercept) storm$Gender_MFM
       3.1679220       -0.5123354
> confint(storm2)
                     2.5 %    97.5 %
(Intercept)       2.816448 3.5640722
storm$Gender_MFM -1.149166 0.1720959
```

The average female hurricane kills 9.5 more people than male storms. The CI95% does not overlap zero so it appears significant. However, Poisson assumes equal variance and mean and that does not appear to match this data, so a negative binomial may be a better option, and when you run a negative binomial you can see that the CI95% crosses zero showing that there is not a significant difference between male and female named storms.

Question 4: My data

```
> soilC<-read.csv("L:/Home/S243/Lab Members/Megan Kelly Slatten/School/mod st
ats/myCdata.csv")
> str(soilC)
'data.frame':   21 obs. of  2 variables:
 $ Species     : Factor w/ 2 levels "BB","SG": 1 1 1 1 1 1 1 1 1 1 ...
 $ Root.number: int  35 27 55 65 45 43 87 67 78 68 ...
> boxplot(soilC$Root.number~soilC$Species)
```




```
> rootmod<-glm(soilC$Root.number~soilC$Species, family="poisson")
> coef(rootmod)
    (Intercept) soilC$SpeciesSG
     4.00884719     -0.04065391
> exp(4.0088-0.04065)-exp(4.0088)
[1] -2.194134
> confint(rootmod)
                    2.5 %     97.5 %
(Intercept)      3.931631 4.08412301
soilC$SpeciesSG -0.158887 0.07682483
```

On average, the big bluestem (BB) species has 2 more roots than the switchgrass (SG) species. However, the CI95% overlaps zero, so there does not appear to be a significant difference in roots.