HW3

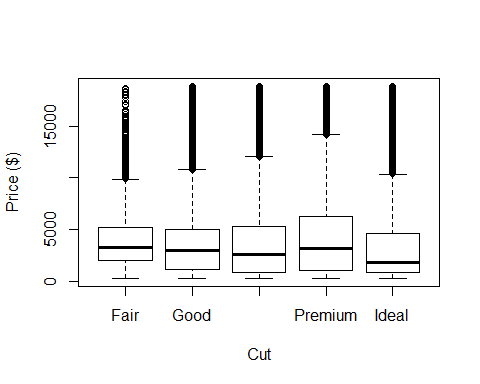
Louis Jochems

February 15, 2019

####Question 1####  
setwd("~/R/homework3-master")  
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

diamonds$cut2<-factor(diamonds$cut,levels=c("Premium","Fair","Good","Very Good","Ideal"))  
diamonds$cut3<-factor(diamonds$cut,levels=c("Fair","Good","Very Good","Premium","Ideal"))  
plot(price~cut3,data=diamonds,xlab="Cut",ylab="Price ($)")



#using a poisson on positive continuous response var   
m1<- glm(price~cut2,data=diamonds,family="poisson")  
coef(m1)

## (Intercept) cut2Fair cut2Good cut2Very Good cut2Ideal   
## 8.4303835 -0.0504411 -0.1542778 -0.1409043 -0.2820703

confint(m1)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 8.43013697 8.43062996  
## cut2Fair -0.05122103 -0.04966133  
## cut2Good -0.15478774 -0.15376782  
## cut2Very Good -0.14127928 -0.14052931  
## cut2Ideal -0.28240541 -0.28173514

#a  
exp(8.4303835) # did premium as baseline because it has has highest median in data

## [1] 4584.258

#b effects of different cuts compared to premium  
##fair  
exp(-0.0504411) #5% decrease OR

## [1] 0.9508099

exp(8.4303835)-exp(8.4303835-0.0504411) #decrease of $237.17

## [1] 225.5

##good  
exp(-0.1542778) #15% decrease OR

## [1] 0.8570339

exp(8.4303835)-exp(8.4303835-0.1542778) #decrease of $746.72

## [1] 655.3934

##very good  
exp(-0.1409043) # 14% decrease OR

## [1] 0.8685724

exp(8.4303835)-exp(8.4303835-0.1409043) # decrease of $693.66

## [1] 602.4979

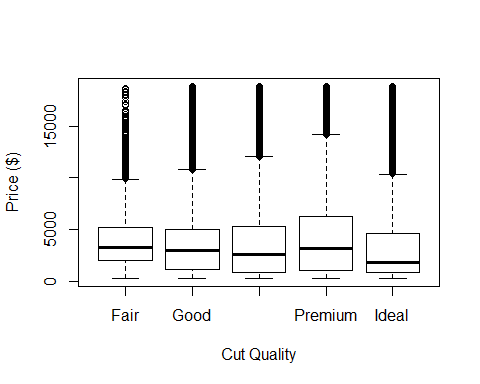
##ideal  
exp(-0.2820703)#25% decrease OR

## [1] 0.7542207

exp(8.4303835)-exp(8.4303835-0.2820703) # decrease of $1493.881

## [1] 1126.716

###OutPut/Parameter Estimates/CI   
plot(price~cut3,data=diamonds,xlab="Cut Quality",ylab="Price ($)")



# With the premium cut as the "control treatment", its coefficient in the model is a baseline price of $4584.26. Comparing all of the other cuts to this price, we see that the cost of fair cut diamonds is about $237.14 less than premium, good cut diamonds cost $746.72 less, very good cuts cost $693.66 less, and the largest decrease of $1493.88 in price for ideal cut. For 95% CI, the effects of all of these cuts compared to premium are significant and negative (all CI's are in negative range and thus do not cross zero).

###Question 2  
setwd("~/R/homework3-master")  
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

contra <- cbind(cuse$using, cuse$notUsing)  
m1<-glm(contra~cuse$education,family="binomial")  
confint(m1)

## Waiting for profiling to be done...

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

coef(m1)

## (Intercept) cuse$educationlow   
## -0.81020374 0.09248529

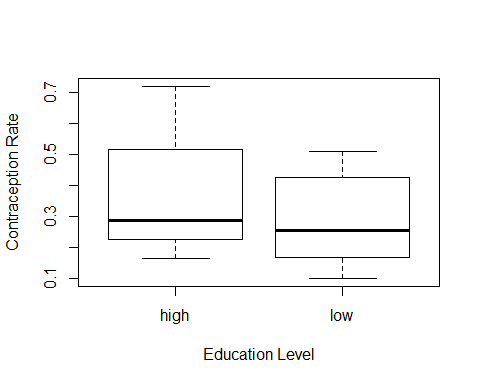
#a  
plogis(-0.81020374)

## [1] 0.3078471

#low from high ed  
plogis(-0.81020374)-plogis(-0.81020374+0.09248529)

## [1] -0.02004851

#Output of Plot/Parameter/CI  
#plotting data as proportion of women using contra out of each obs sample size   
Contraception\_Rate<-cuse$using/cuse$Total  
plot(Contraception\_Rate~cuse$education, ylab="Contraception Rate", xlab="Education Level")

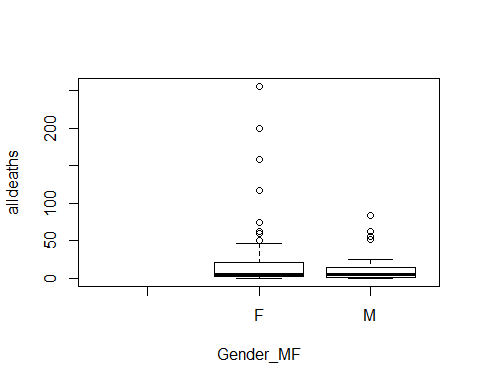


# The coefficient, or baseline, rate of contraception is around 30% among women with low education levels in this study. The effect of education level is low, with only a 0.2% difference between the two categories. The CI's for education crosses zero, indicating there is a pretty good chance of zero effect of education level on contraception (along with a non significant p value of 0.4). Therefore, I accept the null hypothesis that there is no effect of education level on contraception rate among married women in this study.

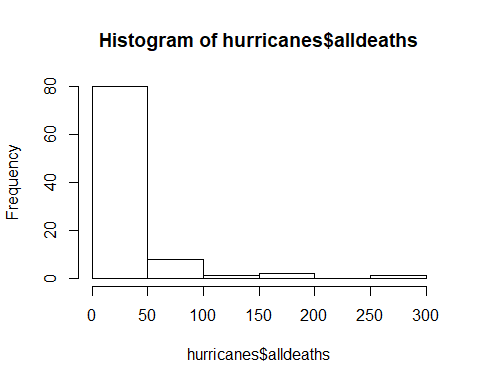
#### Question 3 ####   
setwd("~/R/homework3-master")  
hurricanes<-read.csv("Hurricane Dataset.csv")   
hurricanes<-hurricanes[!apply(hurricanes == "",1, all),]  
head(hurricanes)

## 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

plot(alldeaths~Gender\_MF, data=hurricanes)



hist(hurricanes$alldeaths)



m1<- glm(hurricanes$alldeaths~hurricanes$Gender\_MF,family="poisson")  
coef(m1)

## (Intercept) hurricanes$Gender\_MFM   
## 3.1679220 -0.5123354

#a  
exp(3.1679220)

## [1] 23.75806

#b  
exp(3.1679220)-exp(3.1679220-0.5123354) # 9.5 less deaths?

## [1] 9.524731

confint(m1)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 3.1164152 3.2185581  
## hurricanes$Gender\_MFM -0.6211542 -0.4056501

library("MASS")

## Warning: package 'MASS' was built under R version 3.4.4

m1<- glm.nb(hurricanes$alldeaths~hurricanes$Gender\_MF)  
summary(m1)

##   
## Call:  
## glm.nb(formula = hurricanes$alldeaths ~ hurricanes$Gender\_MF,   
## init.theta = 0.4552963986, link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.90222 -1.15740 -0.74424 -0.05072 2.58074   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.1679 0.1900 16.672 <2e-16 \*\*\*  
## hurricanes$Gender\_MFM -0.5123 0.3342 -1.533 0.125   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(0.4553) family taken to be 1)  
##   
## Null deviance: 111.29 on 91 degrees of freedom  
## Residual deviance: 109.08 on 90 degrees of freedom  
## (6 observations deleted due to missingness)  
## AIC: 708.47  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.4553   
## Std. Err.: 0.0626   
##   
## 2 x log-likelihood: -702.4720

coef(m1)

## (Intercept) hurricanes$Gender\_MFM   
## 3.1679220 -0.5123354

confint(m1)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 2.816448 3.5640722  
## hurricanes$Gender\_MFM -1.149166 0.1720959

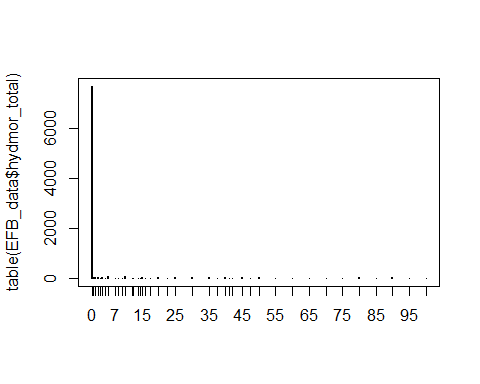
## Since these data are overdispersed (see historgram of deaths), perhaps Jung et al. could have instead used a negative binomial model to more optimally estimate these data. The negative binomial glm includes k (Theta), the overdispersion parameter, which allows the model to treat more of the variance of these data due to random chance and not solely from the effect of gender name of hurricanes. Under this model, we see that while we still have the same coefficient (9.5 more deaths from female name compared to male names), we no longer have a significant effect of gender name on the number of deaths (CI interval crosses zero and p value now 0.12).

#### Question 4 ####   
setwd("~/R/homework3-master")  
EFB\_data<-read.csv("data14\_15\_EFB.csv")  
head(EFB\_data)

## X site name transect\_num point\_num zone Year  
## 1 1 1025 East Sleeping River Wetland 1 1 Meadow 2014  
## 2 2 1025 East Sleeping River Wetland 1 10 Emergent 2014  
## 3 3 1025 East Sleeping River Wetland 1 2 Meadow 2014  
## 4 4 1025 East Sleeping River Wetland 1 3 Meadow 2014  
## 5 5 1025 East Sleeping River Wetland 1 4 Meadow 2014  
## 6 6 1025 East Sleeping River Wetland 1 5 Meadow 2014  
## lat lon lakeCAT lake Geomorph.Subtype NewGeoCat fetch  
## 1 46.984 -89.08742 4 LS Drowned River-mouth 3 18  
## 2 46.984 -89.08742 4 LS Drowned River-mouth 3 18  
## 3 46.984 -89.08742 4 LS Drowned River-mouth 3 18  
## 4 46.984 -89.08742 4 LS Drowned River-mouth 3 18  
## 5 46.984 -89.08742 4 LS Drowned River-mouth 3 18  
## 6 46.984 -89.08742 4 LS Drowned River-mouth 3 18  
## water\_depth org\_depth detritus standing\_dead Grand\_total typ\_pres  
## 1 21 100 2 1 87 0  
## 2 80 100 1 4 25 0  
## 3 28 100 3 2 85 0  
## 4 28 100 3 6 83 0  
## 5 26 100 20 10 85 0  
## 6 42 100 10 10 50 0  
## typ\_total phaaru\_pres phaaru\_total phraus\_pres phraus\_total hydmor\_pres  
## 1 0 0 0 0 0 0  
## 2 0 0 0 0 0 0  
## 3 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 5 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## hydmor\_total ontario erie otherlake emergent submergent meadow  
## 1 0 0 0 1 0 0 1  
## 2 0 0 0 1 1 0 0  
## 3 0 0 0 1 0 0 1  
## 4 0 0 0 1 0 0 1  
## 5 0 0 0 1 0 0 1  
## 6 0 0 0 1 0 0 1

plot(table(EFB\_data$hydmor\_total))  
  
# My response variables, hydmor\_total is the total percent cover of Frogbit in each quadrat, and hydmor\_pres is binomial whether it is absent (0) or present (1) in the transect plots. Therefore, I can run a beta regression model on the positive continuous % cover data and a binomial glm on the presence data. I chose water\_depth is my sole predictor variable to see whether it significantly affects frogbit presence/abundance in Michigan wetlands.   
  
library("betareg")

## Warning: package 'betareg' was built under R version 3.4.4



library("ggplot2")

## Warning: package 'ggplot2' was built under R version 3.4.4

##   
## Attaching package: 'ggplot2'

## The following object is masked \_by\_ '.GlobalEnv':  
##   
## diamonds

EFB\_data$EFBpropveg<- EFB\_data$hydmor\_total/100  
# removing all 0 and 1 values so that the beta reg is able to fit these data.   
beta\_dat1<-EFB\_data[-which(EFB\_data$EFBpropveg==0),]  
beta\_dat<-beta\_dat1[-which(beta\_dat1$EFBpropveg==1),]  
dim(beta\_dat)

## [1] 633 34

#beta regression  
m1<-betareg(EFBpropveg~water\_depth,data=beta\_dat)  
coef(m1)

## (Intercept) water\_depth (phi)   
## -0.809661449 -0.004357125 2.143498109

plogis(-0.809661449)

## [1] 0.3079626

confint(m1)

## 2.5 % 97.5 %  
## (Intercept) -0.96362891 -0.65569399  
## water\_depth -0.00687023 -0.00184402  
## (phi) 1.92361176 2.36338446

# including quadratic(looks like it may fit these data better)  
m2<-betareg(EFBpropveg~water\_depth+I(water\_depth^2),data=beta\_dat)  
coef(m2)

## (Intercept) water\_depth I(water\_depth^2) (phi)   
## -1.0301773490 0.0058898232 -0.0000805679 2.1706261826

plogis(-1.0301773490)

## [1] 0.2630497

#effect size   
plogis(-1.0301773490+0.0058898232\*250)-plogis(-1.0301773490+0.0058898232\*50)

## [1] 0.2848536

confint(m2)

## 2.5 % 97.5 %  
## (Intercept) -1.2511438128 -8.092109e-01  
## water\_depth -0.0016350250 1.341467e-02  
## I(water\_depth^2) -0.0001344292 -2.670662e-05  
## (phi) 1.9477483294 2.393504e+00

#binomial   
m3<-glm(hydmor\_pres~water\_depth, data=EFB\_data,family="binomial")  
coef(m3)

## (Intercept) water\_depth   
## -2.5038308574 0.0002927884

plogis(-2.5038308574)

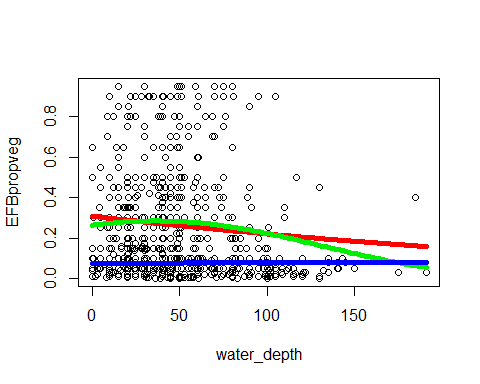
## [1] 0.07559006

confint(m3)

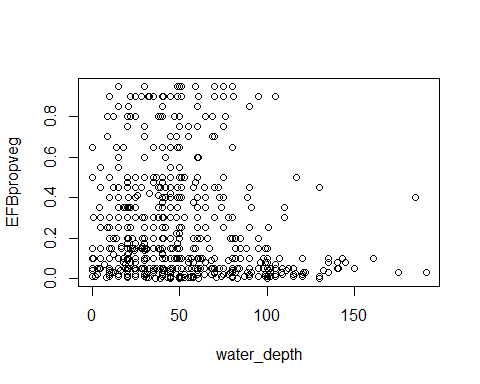
## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -2.624174016 -2.386215851  
## water\_depth -0.001498569 0.002041471

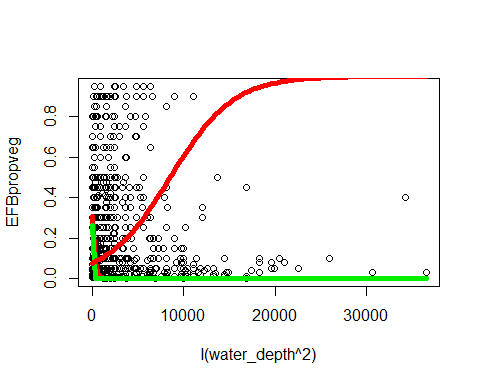
plot(EFBpropveg~water\_depth,data=beta\_dat)  
#linear  
curve(plogis(-0.809661+-0.004537\*x),col="red",lwd=5,add=T)  
#quadratic  
curve(plogis(-1.03018+0.0058898\*x+coef(m2)[3]\*x^2),col="green2",lwd=5,add=T)  
#binom curve  
curve(plogis(-2.50383+0.0002927884\*x),col="blue",lwd=5,add=T)



#changing scale of water\_depth  
plot(EFBpropveg~water\_depth+I(water\_depth^2),data=beta\_dat)



#linear  
curve(plogis(-0.809661+-0.004537\*x),col="red",lwd=5,add=T)  
#quadratic  
curve(plogis(-1.03018+0.0058898\*x+coef(m2)[3]\*x^2),col="green2",lwd=5,add=T)  
#binom curve  
curve(plogis(-2.50383+0.0002927884\*x),col="red",lwd=5,add=T)



### For these analyses, there is a significant effect of water depth on Frogbit percent cover from the beta regressions on both the linear and quadratic models. In particular, with the quadratic model, there is a 28% decrease in EFB percent cover when depth increases from 50 cm to 250 cm. Moreover, the CIs for both the beta linear/quadratic models do not cross zero, indicating a significant effect of water depth on frogbit abundance. In contrast, there was no significant effect of water depth on frogbit presence (CI crosses zero and large p value). Overall, it seems that water depth does not affect whether frogbit is present or absent at a wetland, but it does affect its relative abundance when it is present. It should be noted that I removed the many zeros in the dataset for the beta regression models, so I may need to account for this overdispersion with further zero-inflated/hurdle models on whether water\_depth (along with other covariates in the dataset) affects frogbit abundance.