2. **Exercise 2**

One question in the 1990 General Social Survey asked subjects how many times they had sexual intercourse in the preceding month. Table 1 shows responses classified by gender.

Response	Male	Female	Response	Male	Female	Response	Male	Female
0	65	128	9	2	2	20	7	6
1	11	17	10	24	13	22	0	1
2	13	23	12	6	10	23	0	1
3	14	16	13	3	3	24	1	0
4	26	19	14	0	1	25	1	3
5	13	17	15	3	10	27	0	1
6	15	17	16	3	1	30	3	1
7	7	3	17	0	1	50	1	0
8	21	15	18	0	1	60	1	0

Table 1: Data from the 1990 General Social Survey

(a) Fit a Poisson GLM with log link and a dummy variable for gender (1=males, 0=females) and explain if the model seems appropriate.

```
setwd("G:\\math\\661")
dat<-read.csv("sex.csv")
dat<-data.frame(
      (rep(dat$Response,2)),
    c(dat$Male,dat$Female),
    as.factor(c(rep(1,nrow(dat)),rep(0,nrow(dat)))) )
names(dat)<-c("response","counts","gender")
str(dat)
```

```
## 'data.frame':    54 obs. of  3 variables:
##  $ response: int  0 1 2 3 4 5 6 7 8 9 ...
##  $ counts  : int  65 11 13 14 26 13 15 7 21 2 ...
##  $ gender  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

```
cbind(head(dat) ,tail(dat))
```

```
##   response counts gender response counts gender
## 1        0     65      1       24      0      0
## 2        1     11      1       25      3      0
## 3        2     13      1       27      1      0
## 4        3     14      1       30      1      0
## 5        4     26      1       50      0      0
## 6        5     13      1       60      0      0
```

```
dat.fit<-glm(response ~ gender, family=poisson, weights=counts, data=dat)
summary(dat.fit)
```

```
##
## Call:
## glm(formula = response ~ gender, family = poisson, data = dat,
##     weights = counts)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -33.191    0.000    3.437    6.126   13.430
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.45936    0.02738  53.302  < 2e-16 ***
## gender1      0.30850    0.03822   8.071 6.95e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4050.8  on 44  degrees of freedom
## Residual deviance: 3985.7  on 43  degrees of freedom
## AIC: 5271.3
##
## Number of Fisher Scoring iterations: 6
```

```
tab<-cbind(dat[which(dat$gender == 0),],dat[which(dat$gender == 1 ),-1])
tab<-tab[,c(1,2,4)];tab[19,2]<-sum(tab[19:nrow(tab),2]);
tab[19,3]<-sum(tab[19:nrow(tab),3]);tab<-tab[1:19,]
names(tab)[2:3]<-c("Female","Male")
tab
```

```
##    response Female Male
## 28        0    128   65
## 29        1     17   11
## 30        2     23   13
## 31        3     16   14
## 32        4     19   26
## 33        5     17   13
## 34        6     17   15
## 35        7      3    7
## 36        8     15   21
## 37        9      2    2
## 38       10     13   24
## 39       12     10    6
## 40       13      3    3
## 41       14      1    0
## 42       15     10    3
## 43       16      1    3
## 44       17      1    0
## 45       18      1    0
## 46       20     13   14
```

```
c(sum(tab[,2]),sum(tab[,1]*tab[,2]));    sum(tab[,1]*tab[,2])/sum(tab[,2])
```

```
## [1]  310 1297
```

```
## [1] 4.183871
```

```
sum( tab[,2]*((tab[,1]- 4.183871 )^2) ) / ( sum(tab[,2]) -1)
```

```
## [1] 29.76867
```

```
c(sum(tab[,3]),sum(tab[,1]*tab[,3]));    sum(tab[,1]*tab[,3])/sum(tab[,3])
```

```
## [1]  240 1297
```

```
## [1] 5.404167
```

```
sum( tab[,3]*((tab[,1]- 4.183871 )^2) ) / ( sum(tab[,3]) -1)
```

```
## [1] 31.30203
```

```
1-pchisq(3985.7,43)
```

```
## [1] 0
```

The sample mean for the 1297 women is 4.183871 with a variance of 29.76867. The sample mean for the 1297 men is 5.404167 with a variance of 31.30203. In both groups the sample variances are about 6-7 times the size of the sample means. This is suggesting overdispersion relative to the Poisson. We also see that the model does not give a good fit to the data (p-value ≈ 0).

(b) Interpret the regression coefficient of gender for the model in (a) and provide a 95% Wald confidence interval for the ratio of means for males versus females.

ssssssssssssssss

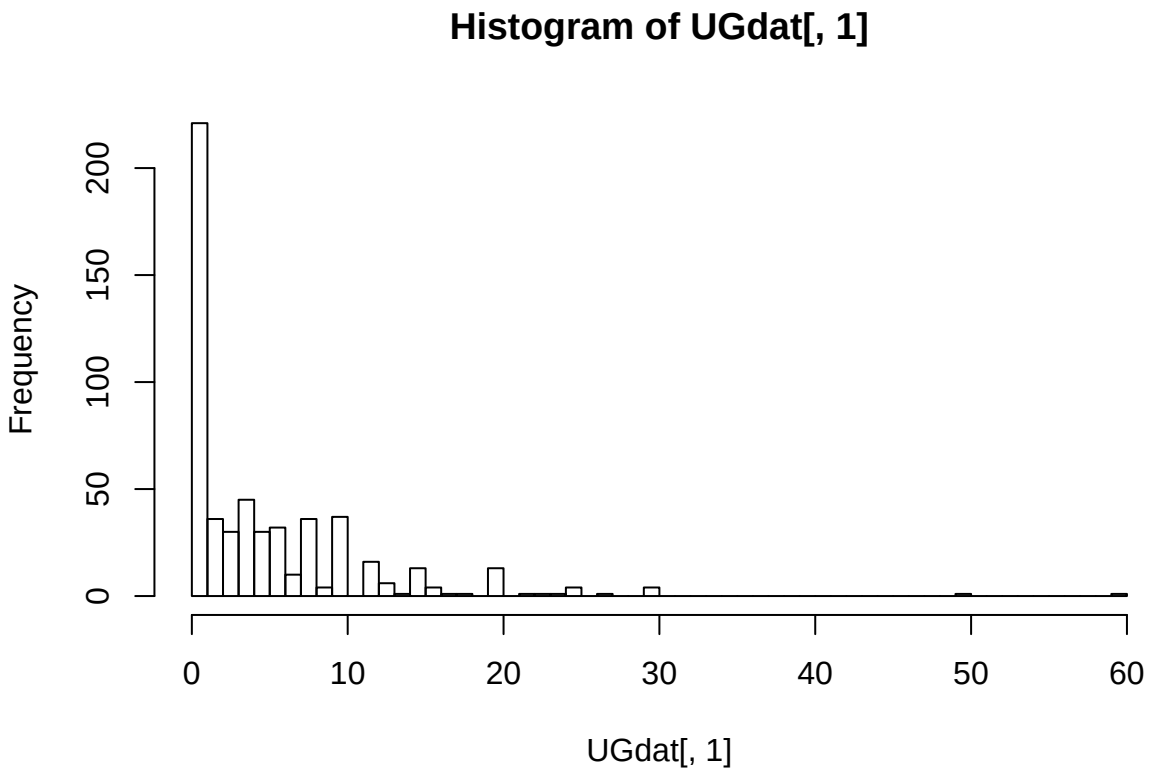
(c) Fit a negative binomial model. Is there evidence of overdispersion? What is the estimated difference in log means, its standard error, and the 95% Wald confidence interval for the ratio of means.

```
library(MASS)
nb.fit<-glm.nb(response ~ gender, weights=counts, data=dat)
summary(nb.fit)
```

```
##
## Call:
## glm.nb(formula = response ~ gender, data = dat, weights = counts,
##     init.theta = 0.5018752366, link = log)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -17.0366    0.0000    0.9873    1.5894    3.4336
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.45936    0.08472  17.226   <2e-16 ***
## gender1      0.30850    0.12724   2.425   0.0153 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.5019) family taken to be 1)
##
##     Null deviance: 606.53  on 44  degrees of freedom
## Residual deviance: 600.60  on 43  degrees of freedom
## AIC: 2883
##
## Number of Fisher Scoring iterations: 1
##
```

```
##
##                Theta:  0.5019
##            Std. Err.:  0.0387
##
##  2 x log-likelihood:  -2876.9770
```

```
UGdat<-as.data.frame(lapply(dat, function(x,p) rep(x,p), dat[["counts"]]))
hist(UGdat[,1], breaks = seq(0,60,by=1))
```



(d) Consider a zero-inflated Poisson model with the zero-inflated component constant across subject (that is with intercept only for the model of ϕ_i). What are the mixing proportions for the degenerate distribution and the Poisson model? Interpret the regression coefficient of gender.

(e) Consider a zero-inflated negative binomial model. What are the mixing proportions for the degenerate distribution and the negative binomial model? Interpret the regression coefficient of gender.

(f) Provide a table with the observed counts and the fitted counts for each of the four models for $y_i = 0, \ldots, 20$ and $y_i > 20$.