HW 10

1. Fit a poisson regression with the variable "doctorco" as the response, and sex, age, income, levyplus, freepoor, freerepa,illness, actdays, and hscore as predicotrs.

```
##
## Call:
## glm(formula = doctorco ~ sex + age + income + levyplus + freepoor +
##     freerepa + illness + actdays + hscore, family = poisson,
##     data = dvisits)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9578  -0.6849  -0.5771  -0.4893   5.6337
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.083736   0.101059 -20.619  < 2e-16 ***
## sex          0.160799   0.055923   2.875  0.00404 **
## age          0.341146   0.162446   2.100  0.03572 *
## income      -0.188581   0.085404  -2.208  0.02724 *
## levyplus     0.139169   0.071291   1.952  0.05092 .
## freepoor    -0.428231   0.179731  -2.383  0.01719 *
## freerepa     0.102124   0.091679   1.114  0.26531
## illness      0.195929   0.017609  11.127  < 2e-16 ***
## actdays      0.127984   0.004907  26.082  < 2e-16 ***
## hscore       0.032181   0.009971   3.227  0.00125 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5634.8  on 5189  degrees of freedom
## Residual deviance: 4384.2  on 5180  degrees of freedom
## AIC: 6735.8
##
## Number of Fisher Scoring iterations: 6
```

2. The expected number of doctor visits for a woman age 50 with an income of 10,000 Australian dollars, not covered by private insurance, provided coverage by the government due to lowincome (but not due to old age), with 2 ilnesses in the past 2 weeks,no days of reduced activity, and a health score of 1 is **0.1467**

```
# make a dataframe with new data
newdata = data.frame(sex=1, age=0.52, # midpoint(50-54)/100
                     income=0.90005,# midpoint(8001-10000)/10000
                     levyplus=0, freepoor=1, freerepa=0, illness= 2, actdays=0, hscore=1)

# Prediction
predict(model, newdata = newdata, type='response')
```

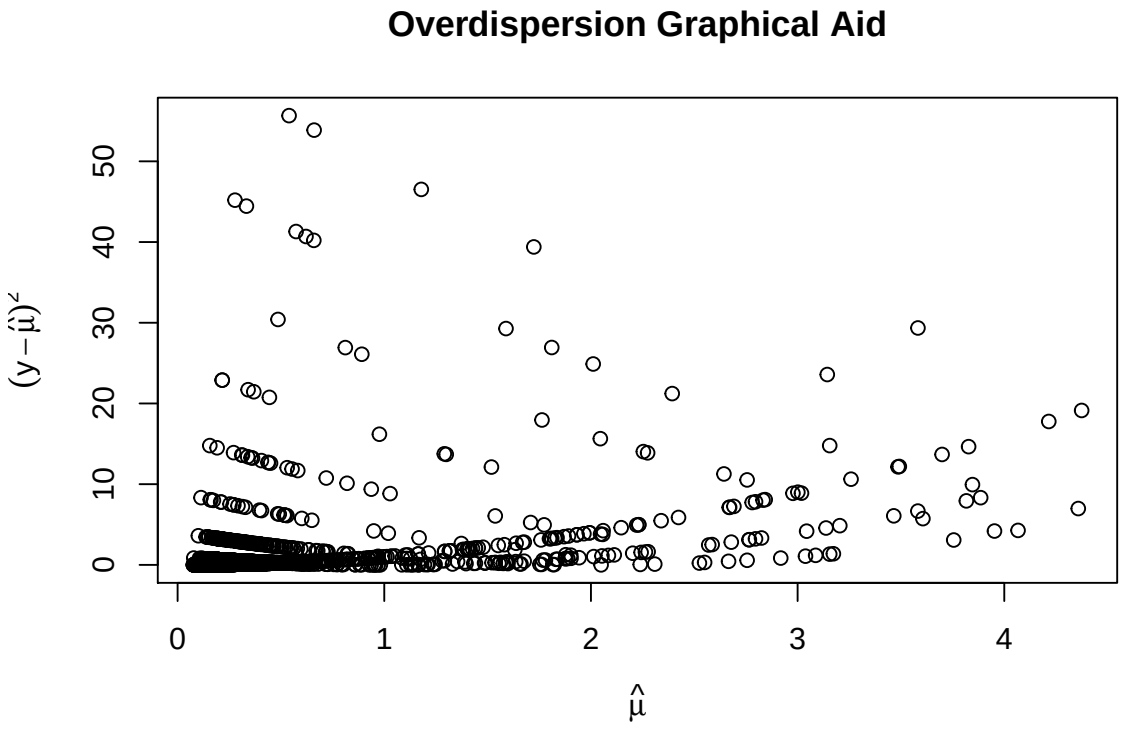
```
##         1
## 0.1466855
```

3. For the person described above, the chance she will have 0 doctor visits is **86.36%**.

```
dpois(x=0, lambda=0.1466855 )
```

```
## [1] 0.8635655
```

4. A Poisson model assumes that the mean structure equals to variance structure. If we allow for a dispersion parameter,in a perfect seceario (mean=variance), the dispersion parametere would equal 1. This assumption is broken when the dispersion parameter is greater/smaller than 1. In our case, the dispersion parameter is **1.33** suggesting **mild overdispersion** (variance greater than mean), which can be confirmed by the graphical aid. Possible reason for overdispersion could be the lack of homogeneity (e.g. *heterogeneity* where subjects within each covariate- **young vs old** combination differ greatly), and/or independence. Also, the problem of overdispersion may be confounded with the problem of omitted covariates- not all available variables have been used in this analysis; the lack of those covariates from the model could be the reason for overdispersion. Finally, the main problem I have been able to discover is the presence of multiple **outliers**; these observations have have high studentized residuals (>3.5) that allows us to reject the null hypothesis (not an outlier) with a level of alpha/n.




```
#Numerical
sigma2= sum(residuals(model, type='pearson')^2)/model$df.residual
sigma2
```

```
## [1] 1.327031
```

After dropping the first outlier of the data, and redoing the analysis, the overdispersion parameter is indeed smaller.

```r
#Sutdentized Residuals
stud <- rstudent(model)

#Ho: not an outlier Ha: outlier
2*(1-pt(max(abs(stud)), model$df.residual-1)) >  (0.05/5190)
```

```
## [1] FALSE
```

```r
#New data without outlier
myData <- dvisits[-c(334), ]

#New Model
model2= glm(doctorco ~ sex+age+income+levyplus+freepoor+freerepa+illness+actdays+ hscore,
            family = poisson, data=myData)

#Smaller Overdispersion
sigma3= sum(residuals(model2, type='pearson')^2)/model2$df.residual
sigma3
```

```
## [1] 1.302058
```