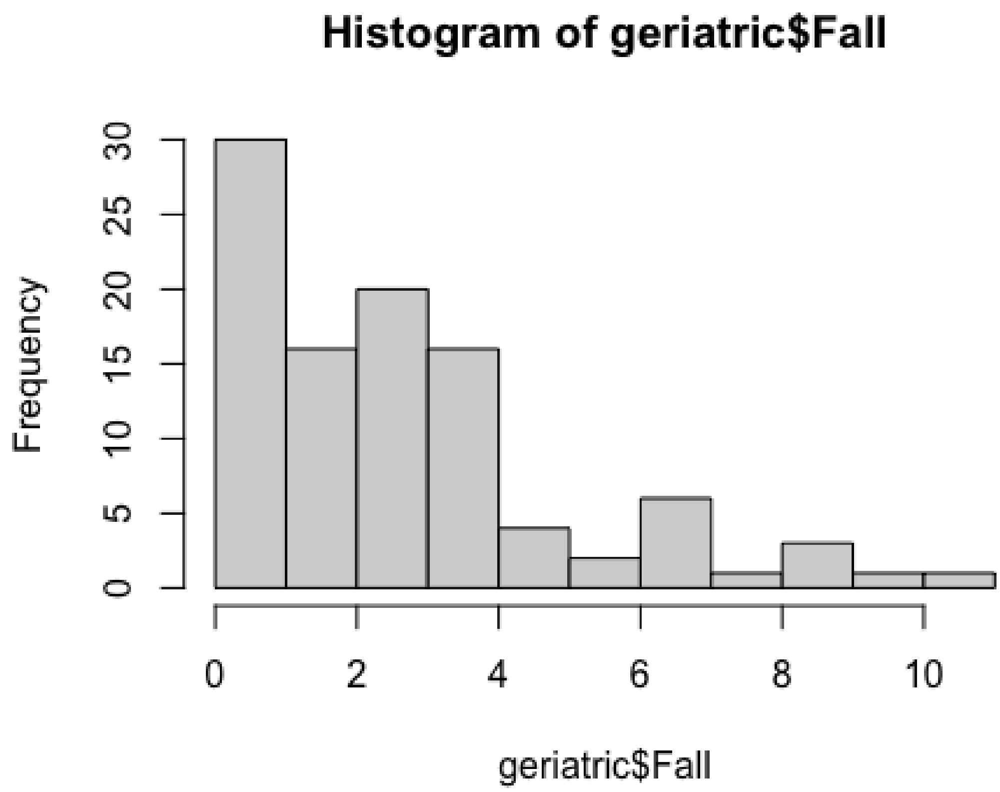
Assignment6_stat359

Koki Itagaki

2023-04-03

```
#A geriatrics researcher studied the effects of two interventions on the
frequency of falls. Subjects were at least 65 years of age and in reasonably
good health. The variables found in the dataset geriatric.txt are: number of
falls, intervention (0=education, 1=education and aerobics), gender
(0=female), balance index, strength index.

#(a) Fit a Poisson regression model based on u = exp(b0+ b1x1 + b2x2 + b3x3+
B4x4). Produce
#a table of estimated coefficients, their estimated standard errors, and the
corresponding
#confidence intervals.
geriatric<-read.table(file='~/Desktop/stat359/data/geriatric.txt', header =
T)
head(geriatric)

##   Fall Int Sex BI SI
## 1    1   1   0 45 70
## 2    1   1   0 62 66
## 3    2   1   1 43 64
## 4    0   1   1 76 48
## 5    2   1   0 51 72
## 6    1   1   1 73 39

table(geriatric$Fall)

##
##  0  1  2  3  4  5  6  7  8  9 10 11
## 12 18 16 20 16  4  2  6  1  3  1  1

hist(geriatric$Fall)
```




```
fit1<-glm(Fall~ Int + Sex + BI + SI,family = poisson,data = geriatric)
summary(fit1)

##
## Call:
## glm(formula = Fall ~ Int + Sex + BI + SI, family = poisson, data =
geriatric)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1854  -0.7819  -0.2564   0.5449   2.3626
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.489467   0.336869   1.453  0.14623
## Int         -1.069403   0.133154  -8.031 9.64e-16 ***
## Sex         -0.046606   0.119970  -0.388  0.69766
## BI           0.009470   0.002953   3.207  0.00134 **
## SI           0.008566   0.004312   1.986  0.04698 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

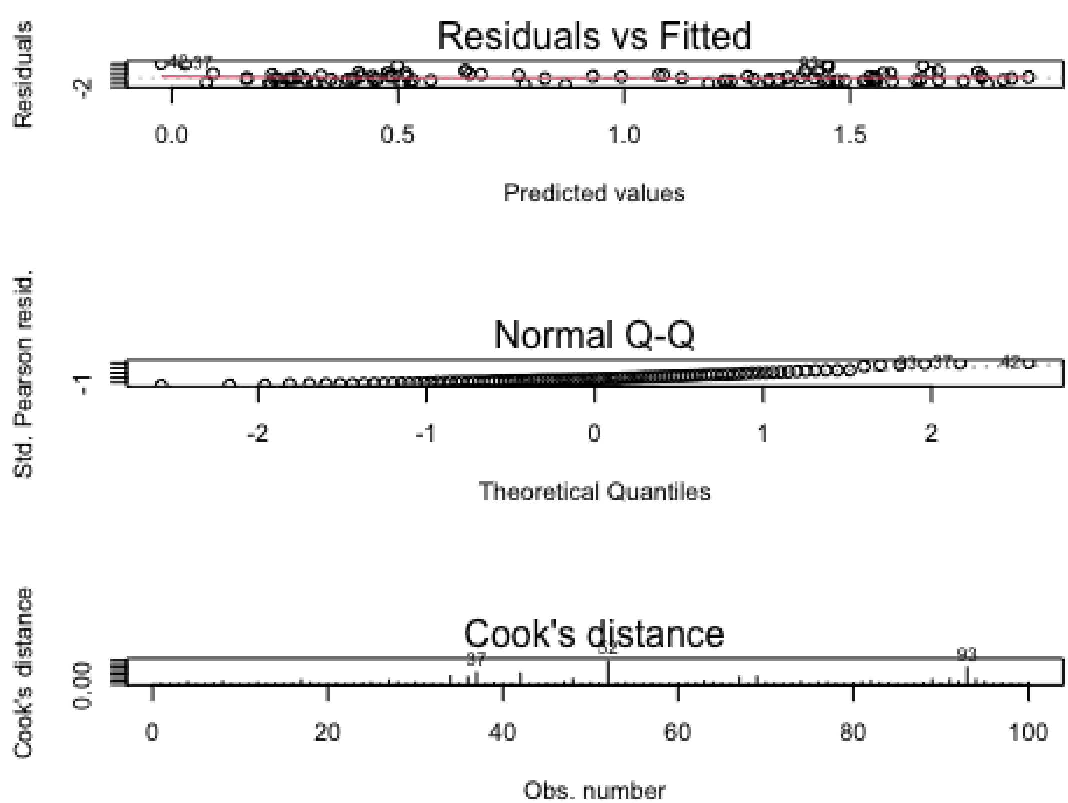
```
##
##     Null deviance: 199.19  on 99  degrees of freedom
## Residual deviance: 108.79  on 95  degrees of freedom
## AIC: 377.29
##
## Number of Fisher Scoring iterations: 5
```

```
confint(fit1)
```

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

```
##                      2.5 %      97.5 %
## (Intercept) -0.1836076944  1.13605432
## Int         -1.3360219299 -0.81332114
## Sex         -0.2823288477  0.18838553
## BI           0.0036833502  0.01526299
## SI           0.0001457923  0.01704817
```

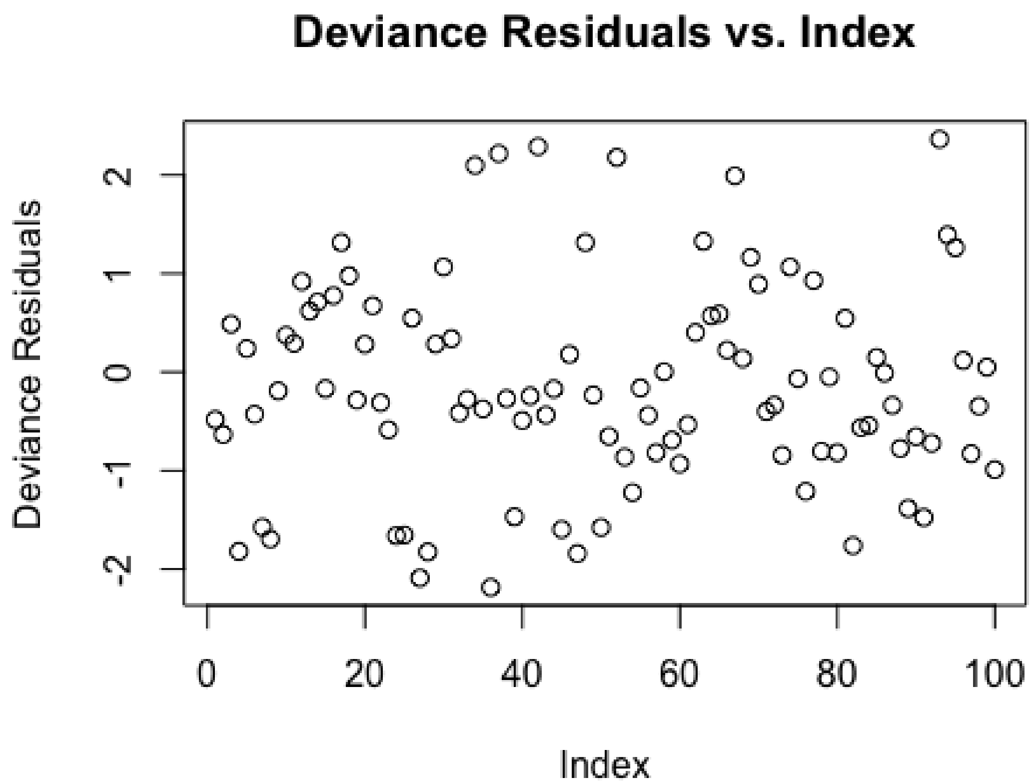
```
par(mfrow = c(3,1))
plot(fit1,which = c(1,2,4))
```




```
#The Residuals vs fitted graph shows almost the constant variance:However, the
#q-q plot shows the distribution is right skewed.
```

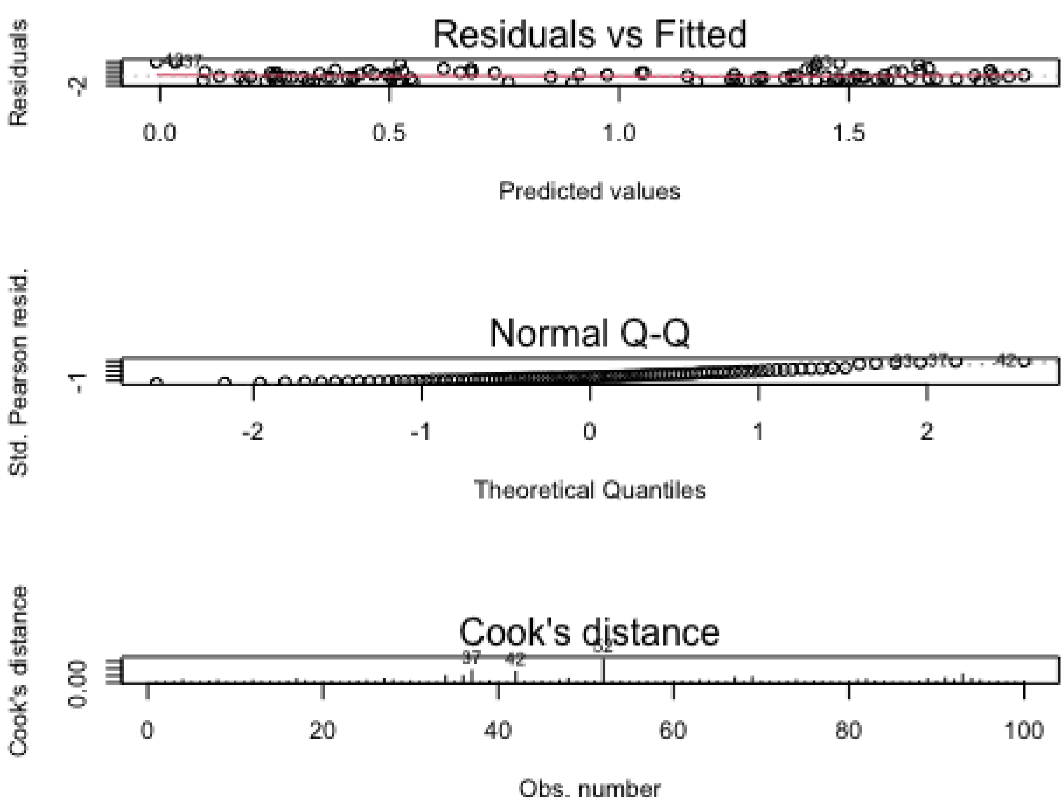
```r
#a table of estimated coefficients, their estimated standard errors, and the corresponding
#confidence intervals.
df<-data.frame(estimate = c(0.489467,-1.069403,-0.046606,0.029104,0.008566),
               est_sderror = c(
  0.336869,0.133154,0.119970,0.002953,0.004312),ci = confint(fit1)
)

## Waiting for profiling to be done...

df

##                estimate est_sderror      ci.2.5..   ci.97.5..
## (Intercept)  0.489467    0.336869 -0.1836076944  1.13605432
## Int         -1.069403    0.133154 -1.3360219299 -0.81332114
## Sex         -0.046606    0.119970 -0.2823288477  0.18838553
## BI           0.029104    0.002953  0.0036833502  0.01526299
## SI           0.008566    0.004312  0.0001457923  0.01704817

#(b) Obtain the model deviance and perform a goodness-of-fit test. State your
conclusion.

deviance_fit1<-deviance(fit1)
deviance_fit1

## [1] 108.7899

res_df_fit1<-(fit1$df.residual)
res_df_fit1

## [1] 95

1-pchisq(deviance_fit1,res_df_fit1)

## [1] 0.157792

#Ho:Model is adequate, Ha: Model not adequate.
#Dispersion parameter for gaussian family taken to be 1.
#But,model deviance is 108.79 and degreees of freedom is 95.
#THese numbers are close. so My conclusion is the model isvgood fit.
#Also, according to the chi-square test, the p-value > 0.05, this means that
#we fail to reject Ho and there is a insignificant evidence that the model
#is not adequate.

#(c) Plot the deviance residuals (versus their index). Do there appear to be
any outlying cases?
dev_res <- residuals.glm(fit1,"deviance")
plot(seq_along(dev_res),dev_res,
     main="Deviance Residuals vs. Index",
```

```
    xlab="Index",
    ylab="Deviance Residuals")
```




```
#From the graph, almost all of the data are gathered around 0 at deviance residuals.
#However there are some outliers whose deviance residuals are approximately 6 and -4.

#(d) Use a deviance test to test the hypothesis that gender can be dropped from the model. What is your conclusion?
#I compute the differrence in deviance as a test statistic and compute the p-value:
#Ho: b2 = 0 Ha: b2 != 0
#remove Sex and the fit the smaller model
fit2<- glm(Fall~ Int + BI + SI,data = geriatric, family = poisson)
#assess the significance of Sex using a deviance test
anova(fit2,fit1,test = "Chi")

## Analysis of Deviance Table
##
## Model 1: Fall ~ Int + BI + SI
## Model 2: Fall ~ Int + Sex + BI + SI
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1        96      108.94
## 2        95      108.79  1     0.151   0.6976

1-pchisq(deviance(fit2)-deviance(fit1), df = 1)

## [1] 0.6975832

#Since p-value > a = 0.05, Based on the likelihood ratio we would not reject
Ho.
#Therefore there is a insignificant evidence that the gender is related to
the number
#of Fall.
par(mfrow = c(3,1))
plot(fit2,which = c(1,2,4))
```




```
#(e) Fit a model without ‘gender’. Obtain an approximate 95% confidence
interval for §; (thecoefficient for intervention’) and interpret the
confidence interval.
fit3<-update(fit1,.~. -Sex)
summary(fit3)

##
## Call:
## glm(formula = Fall ~ Int + BI + SI, family = poisson, data = geriatric)
```

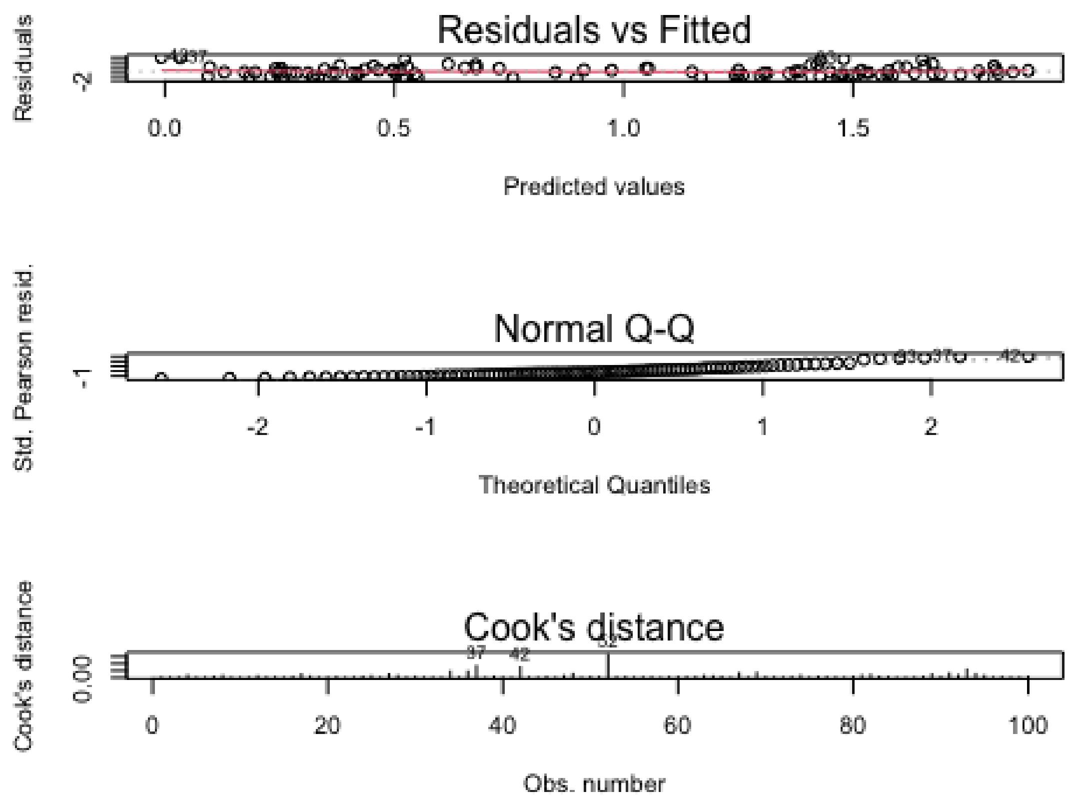
```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2152  -0.7512  -0.2594   0.5830   2.2893
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.443890   0.317289   1.399  0.16181
## Int         -1.077770   0.131415  -8.201 2.38e-16 ***
## BI           0.009471   0.002957   3.203  0.00136 **
## SI           0.008979   0.004190   2.143  0.03209 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 199.19  on 99  degrees of freedom
## Residual deviance: 108.94  on 96  degrees of freedom
## AIC: 375.44
##
## Number of Fisher Scoring iterations: 5
```

```
confint(fit3, "Int")
```

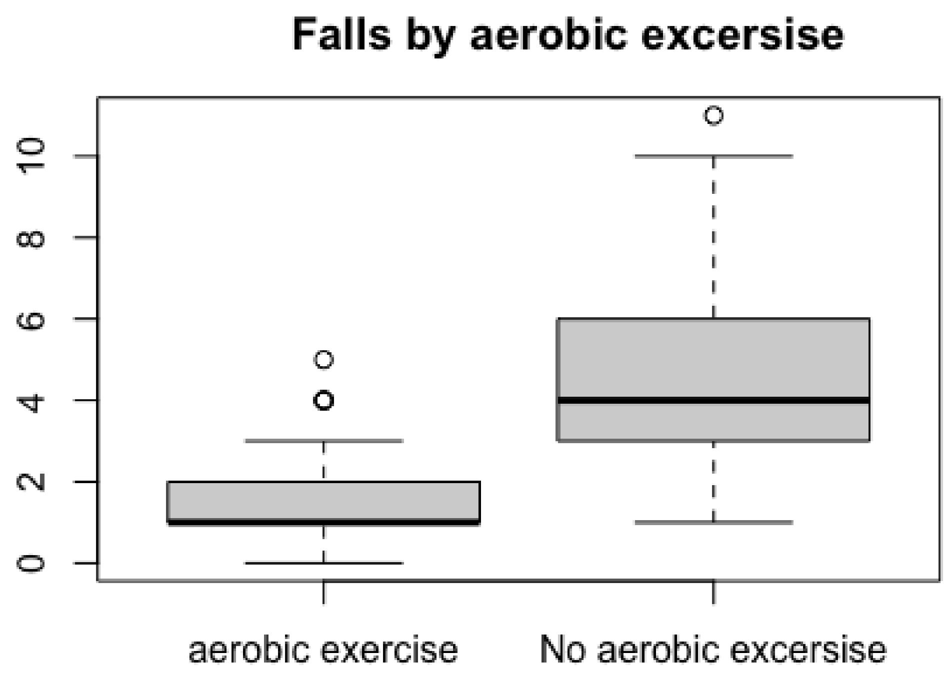
```
## Waiting for profiling to be done...
```

```
##      2.5 %     97.5 %
## -1.3411097 -0.8252039
```

```
par(mfrow = c(3,1))
plot(fit2,which = c(1,2,4))
```




```
# The 95% confidence interval for the intervention is between -3.7198 and -
2.2801.
#the range of the interval is lower than 0. This means that there is a
negative relationships between Fall and intervention. So, If the more the
intervention is,
#the more the number of falls decreases.

#(f) Is aerobic exercise associated with a reduction in the frequency of
falls
#when controlling for balance and strength?

boxplot(geriatric$Fall[geriatric$Int == 1], geriatric$Fall[geriatric$Int ==
0],
        names = c("aerobic exercise", "No aerobic excersise"), main = "
        Falls by aerobic excersise")
```




```
#I just put the box plot to show how Int affects the frequency of Fall.
#According to this graph, it shows that Int might cause a decrease in Fall.



#Ho: B(int) = 0
#Ha: B(int) != 0
fit3<-glm(Fall~ Int + BI + SI,data = geriatric,family = poisson)
fit4<-glm(Fall~ BI + SI,data = geriatric,family = poisson)

summary(fit3)

##
## Call:
## glm(formula = Fall ~ Int + BI + SI, family = poisson, data = geriatric)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2152  -0.7512  -0.2594   0.5830   2.2893
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.443890   0.317289   1.399  0.16181
```

```
## Int         -1.077770   0.131415  -8.201 2.38e-16 ***
## BI           0.009471   0.002957   3.203  0.00136 **
## SI           0.008979   0.004190   2.143  0.03209 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 199.19  on 99  degrees of freedom
## Residual deviance: 108.94  on 96  degrees of freedom
## AIC: 375.44
##
## Number of Fisher Scoring iterations: 5
```

```
anova(fit3,fit4,test = "Chi")
```

```
## Analysis of Deviance Table
##
## Model 1: Fall ~ Int + BI + SI
## Model 2: Fall ~ BI + SI
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1        96     108.94
## 2        97     186.10 -1  -77.158 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(fit4)
```

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

```
##                     2.5 %     97.5 %
## (Intercept) -0.4587608707 0.74083228
## BI           0.0031585624 0.01488697
## SI          -0.0003678417 0.01572167
```

```
#I computed the difference in deviance as a test statistic and compute the p-
value.
#Based on the likelihood ratio,we reject Ho. There is a significant evidence
that
#Int is assorciated with the frequency of Fall.

#Also, according to the summary of the fit3 poisson regression,the p-value is
#1.25e-12 <<0.05, so It is significant that aerobic exercise is assorciated
#with the frequency of falls. Since the estimate of the Int is -2.999965,
also
#the 95% of the confidence interval of Int is between -3.7197929927 -
2.28013697.
#Therefore, aerobic exercise is associated with a reduction in the frequency
of falls
#when controlling for balance and strength.
```