Assignment 4

Zimeng Ming V00844078

2019/3/28

Question1:

First of all, we load the data to the system:

```
geriatric<-read.table("geriatric.txt",header = T)
summary(geriatric)
```

```
##       Fall            Int           Sex              BI       
##  Min.   : 0.00   Min.   :0.0   Min.   :0.00   Min.   :13.00  
##  1st Qu.: 1.00   1st Qu.:0.0   1st Qu.:0.00   1st Qu.:39.00  
##  Median : 3.00   Median :0.5   Median :1.00   Median :51.50  
##  Mean   : 3.04   Mean   :0.5   Mean   :0.53   Mean   :52.83  
##  3rd Qu.: 4.00   3rd Qu.:1.0   3rd Qu.:1.00   3rd Qu.:66.25  
##  Max.   :11.00   Max.   :1.0   Max.   :1.00   Max.   :98.00  
##        SI       
##  Min.   :18.00  
##  1st Qu.:52.00  
##  Median :60.00  
##  Mean   :60.78  
##  3rd Qu.:70.25  
##  Max.   :90.00  
```

Q1: Fit a possion regression model

```
#according to the question:
x1=geriatric$Int
x2=geriatric$Sex
x3=geriatric$BI
x4=geriatric$SI
y=geriatric$Fall

fit1<-glm(y~x1+x2+x3+x4,family = poisson)
summary(fit1)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3 + x4, family = poisson)
##
## 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
## x1          -1.069403   0.133154  -8.031 9.64e-16 ***
## x2          -0.046606   0.119970  -0.388  0.69766
## x3           0.009470   0.002953   3.207  0.00134 **
## x4           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
```

So for Intervention, The Estimate corfficients is -1.069403, The Estimate standard errors is 0.133154

So for Sex, The Estimate corfficients is -0.046606, The Estimate standard errors is 0.119970

So for Balance index, The Estimate corfficients is 0.009470, The Estimate standard errors is 0.002953

So for Strength index, The Estimate corfficients is 0.008566, The Estimate standard errors is 0.004312

The estimate response function u=exp(0.489467-1.069403*Intervention - 0.046606*Sex+0.009470*Balance Index +0.008566*Strength index)

---

b):

```
#According the summary from a), we know the residual deviance is 108.79, the degree o
f freedom is 95
#To calculate the p-value for the deviance goodness of fit test we need to calculate
the p-value for the chi-squared distribution.
cat("The model deviance is ", fit1$deviance,'\n')
```

```
## The model deviance is  108.7899
```

```
pchi_fit1<-1-pchisq(fit1$deviance,df=fit1$df.residual,lower.tail = F)
cat("The p-value is",pchi_fit1)
```

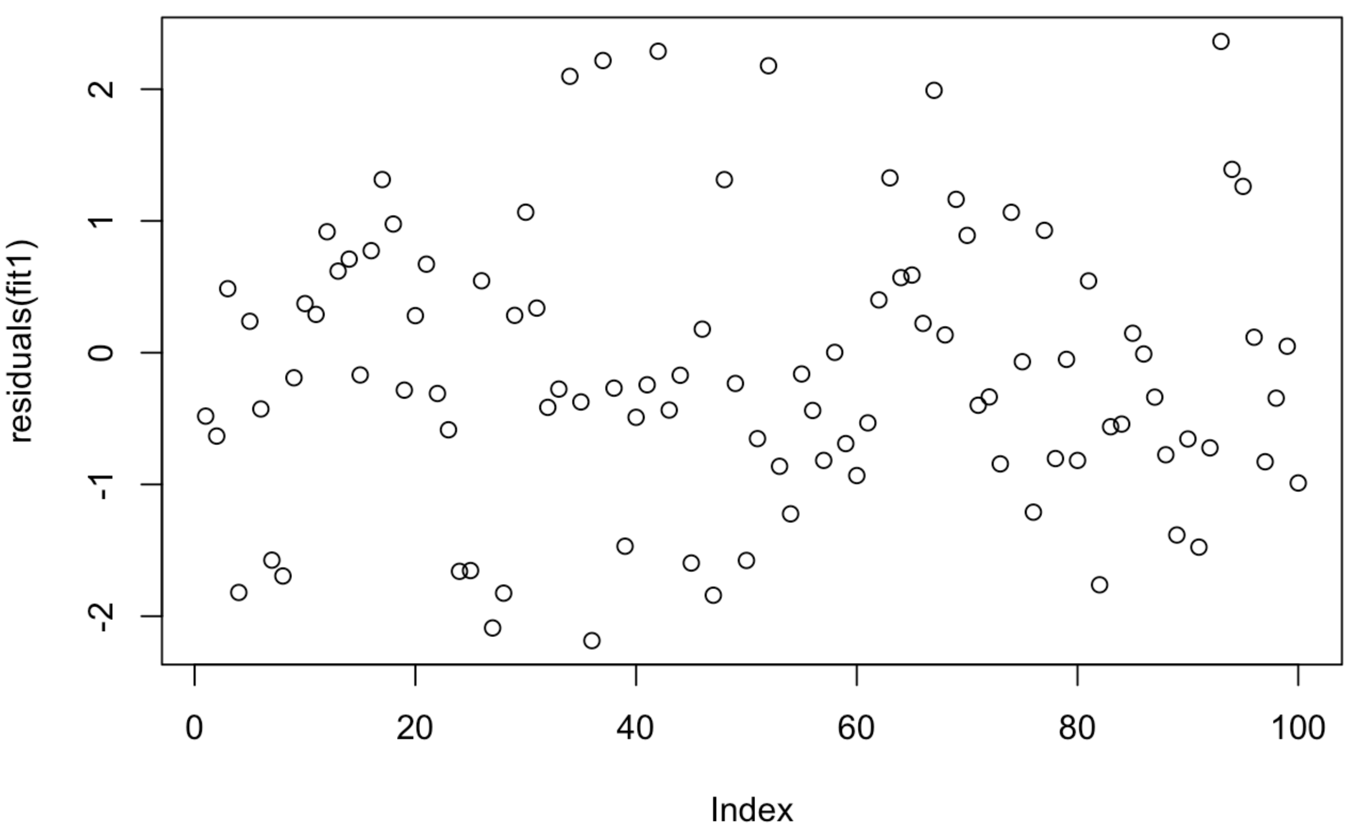
```
## The p-value is 0.842208
```

The null hypothesis is the model is fitted. Since the P-value are large, we have little evidence to aginst null hypothesis. Thus we do not have sufficient evidence to say the model is not fitted.

---

c):

```
plot(residuals(fit1))
```



There is no outliers shows on the graph. It seems looks like all data are normal.

---

d):

```
#so if we want test the hypothesis that sex can be dropped from the model, we first u
pdate the regression model and delete sex from the model

fit2<-update(fit1,~.-x2)
summary(fit2)
```

```
##
## Call:
## glm(formula = y ~ x1 + x3 + x4, family = poisson)
##
## 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
## x1          -1.077770   0.131415  -8.201 2.38e-16 ***
## x3           0.009471   0.002957   3.203  0.00136 **
## x4           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
```

Now for the AIC test:

```
cat("fit1 AIC=", fit1$aic, " deltaAIC=", fit1$aic-min(fit1$aic,fit2$aic),'\n')
```

```
## fit1 AIC= 377.2878  deltaAIC= 1.849002
```

```
cat("fit2 AIC=", fit2$aic, " deltaAIC=", fit2$aic-min(fit1$aic,fit2$aic))
```

```
## fit2 AIC= 375.4388  deltaAIC= 0
```

for the deivence test:

```
anova(fit2,fit1,test = "Chi")
```

```
## Analysis of Deviance Table
##
## Model 1: y ~ x1 + x3 + x4
## Model 2: y ~ x1 + x2 + x3 + x4
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        96     108.94
## 2        95     108.79  1    0.151   0.6976
```

Here we can see from the two tests above, The deltaAIC for fit2 is 0, and the P-value for the deivence test is very large. There the evidence shows that little eveidence agiainst the hypothessis that the coefficient of sex is 0. So the sex can be dropped from the model.

---

e):

```
#the model without X2 is fit2.
summary(fit2)
```

```
##
## Call:
## glm(formula = y ~ x1 + x3 + x4, family = poisson)
##
## 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
## x1          -1.077770   0.131415  -8.201 2.38e-16 ***
## x3           0.009471   0.002957   3.203  0.00136 **
## x4           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
```

From the summary above, we know the Beta1 is -1.077770 and the Standard Error is 0.131415

```
cat("The 95% confidence interval is", '\n', cbind(-1.077770-1.96*0.131415,-1.077770+1
.96*0.131415))
```

```
## The 95% confidence interval is
##  -1.335343 -0.8201966
```

The interpret confidence interval is 95% and the confidence interval do not contain the 0, so also improve that the sex can be dropped.

---

f): now we consider two fit models. For the first model with the sex, we know from the above question that the sex can be removed from the model, so we use fit2(without sex) model

For the model without Sex

```
exp(fit2$coefficients)
```

```
## (Intercept)          x1          x3          x4
##   1.5587593   0.3403536   1.0095161   1.0090196
```

So the estimated risk is 0.3403536 when the Strength and Balance get controlled, and the aerobic exercise is negatively assocaited with frequency of falls since the correlation is negative(-1.077770).