

Logistic Regression

Sharad Deshmukh

13/08/2021

Introduction

```
D=read.csv(file.choose(),header=TRUE)

D=na.omit(D)  ##ommit the missing value
y=ifelse(D$chd=="no",0,1)
x1=D$height
x2=D$cigs
n=length(y)
lmod=glm(y~x1+x2,family=binomial)
summary(lmod)  #####test model with all parameter
```

To understand the causes of heart disease, 3154 healthy men, aged from 39 to 59, from the SanFrancisco area were studied. Eight and a half years later, the study recorded whether these men now suffered from heart disease along with many other variables that might be related to the chance of developing this disease. The variables include chd' indicating whether the heart disease was present or absent,height' (in inches) and the number of cigarettes (cigs') smoked per day. The data is stored in a file titledwcgs'.

```
##
## Call:
## glm(formula = y ~ x1 + x2, family = binomial)
##
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max
## -1.0161   -0.4421   -0.3597   -0.3488    2.4285
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.219165  1.852850 -2.277  0.0228 *
## x1          0.021019  0.026495  0.793  0.4276
## x2          0.023584  0.004059  5.810 6.24e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1769.2 on 3139 degrees of freedom
## Residual deviance: 1736.3 on 3137 degrees of freedom
```

```

## AIC: 1742.3
##
## Number of Fisher Scoring iterations: 5

```

```

lmod1=glm(y~x2,family=binomial)  ##test with parameter x2
anova(lmod1,lmod,test="Chi")   ##deviance=0 model is perfect

```

Fit the logistic model with parameter x2

```

## Analysis of Deviance Table
##
## Model 1: y ~ x2
## Model 2: y ~ x1 + x2
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      3138    1737.0
## 2      3137    1736.3  1  0.63111  0.4269

```

```

lmodc1=glm(y~x1,family=binomial)
anova(lmodc1,lmod,test="Chi")

```

```

## Analysis of Deviance Table
##
## Model 1: y ~ x1
## Model 2: y ~ x1 + x2
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      3138    1768.3
## 2      3137    1736.3  1  32.011 1.533e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

reg=glm(y~x2,family=binomial(link="logit"),data=D) ##LOGistic Model
summary(reg)

```

```

##
## Call:
## glm(formula = y ~ x2, family = binomial(link = "logit"), data = D)
##
## Deviance Residuals:
##       Min      1Q      Median      3Q      Max
## -1.0094 -0.4415 -0.3516 -0.3516  2.3724
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.75239   0.09266 -29.70 < 2e-16 ***
## x2          0.02367   0.00406   5.83 5.54e-09 ***

```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1769.2 on 3139 degrees of freedom
## Residual deviance: 1737.0 on 3138 degrees of freedom
## AIC: 1741
##
## Number of Fisher Scoring iterations: 5

```

```

fit=reg$fitted.values
summary(fit)

```

```

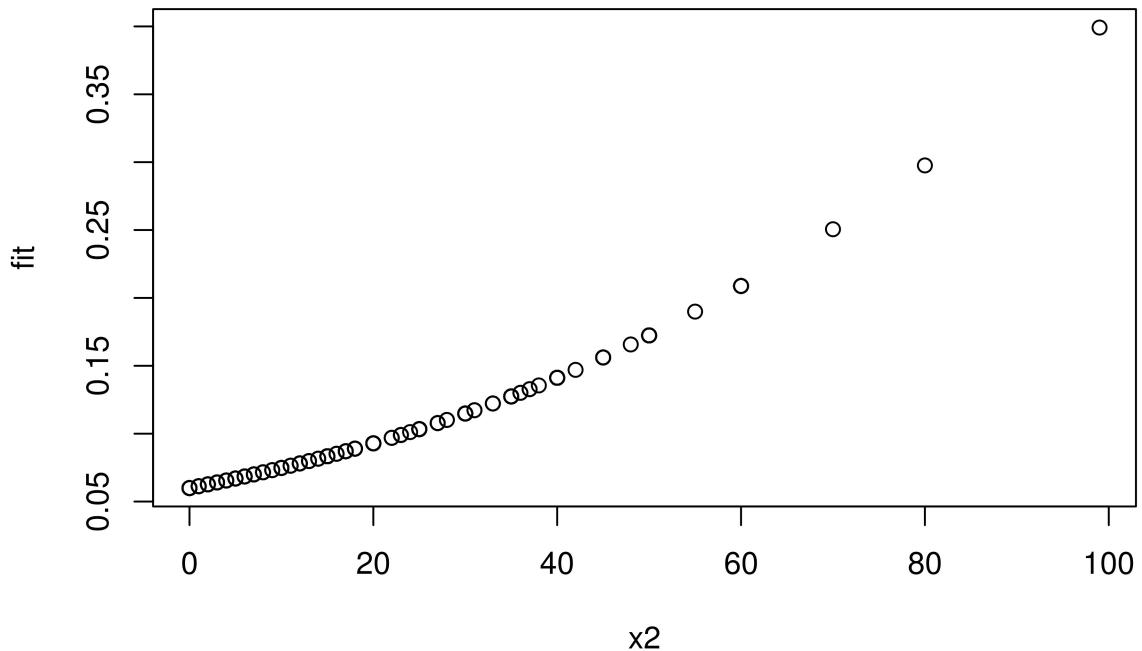
##      Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.05995 0.05995 0.05995 0.08121 0.09288 0.39916

```

```

plot(x2,fit) ## as we increased spoke person the chance CHD incrases```

```



as we increased smoke person the chance CHD increases

```

n1=length(which(y==0));n1

```

```

## [1] 2885

```

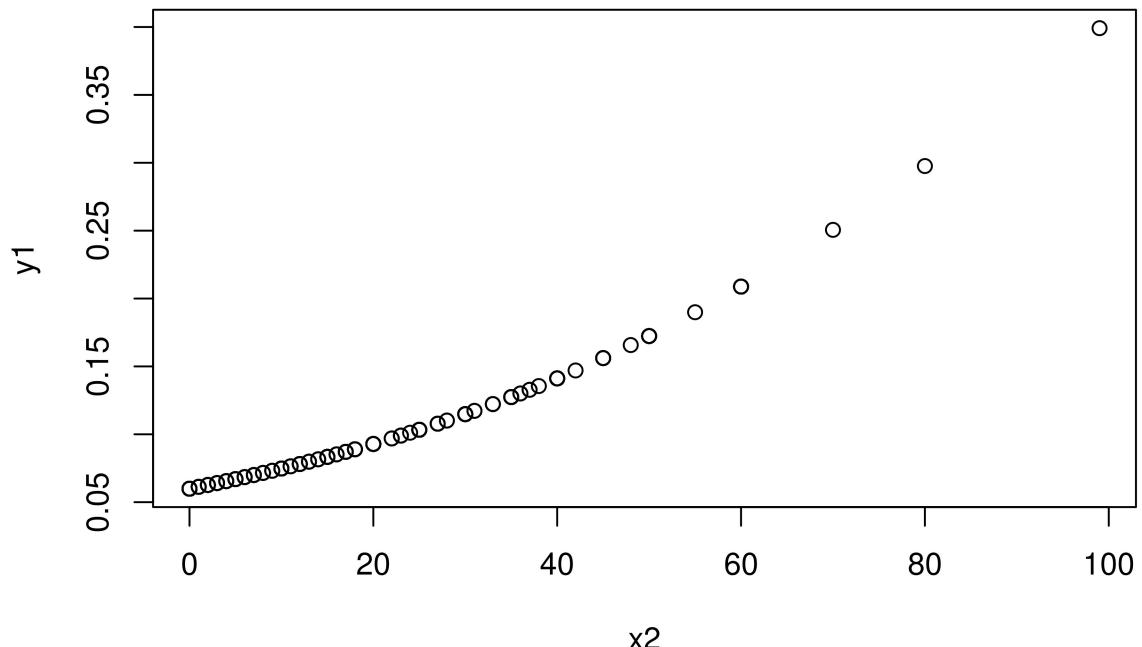
```

 $1-n1/n$ 

## [1] 0.08121019

###fitted model
y1=1/(1+exp(2.7523-0.02367*x2))
plot(x2,y1)

```



```

##odds ratio
exp(0.02367) ##1 cigarettes per day

## [1] 1.023952

##if we increased 20 cigarettes per day
exp(20*0.02367) ##we have 60% chance getting CHD,

```

```
## [1] 1.605443
```

if we increased 20 cigarettes per day then there is 60% chance getting CHD,