

R Notebook

Code ▼

This is an R Markdown (<http://rmarkdown.rstudio.com>) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Hide

```
#question 1
#part a
library(openxlsx)
xlsxFile <- system.file("extdata","hw_9_data.xlsx", package = "openxlsx")
df <- read.xlsx(xlsxFile = "hw_9_data.xlsx", sheet = 1, skipEmptyRows = FALSE)

mdl1= glm( y ~ x, data = df, family = binomial)
#part b
summary(mdl1)
```

Call:

```
glm(formula = y ~ x, family = binomial, data = df)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0620	-0.4868	0.3915	0.5476	2.1682

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.070884	2.108996	2.879	0.00399 **
x	-0.017705	0.006076	-2.914	0.00357 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 34.617 on 24 degrees of freedom
 Residual deviance: 20.364 on 23 degrees of freedom
 AIC: 24.364

Number of Fisher Scoring iterations: 4

Hide

```
#install.packages("ResourceSelection")
library(ResourceSelection)
hoslem.test(df$y,fitted(mdl1), g=8) ##### NOT TESTED
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: df$y, fitted(md11)
X-squared = 4.8315, df = 6, p-value = 0.5656
```

Hide

```
#part c
OR=exp(coef(md11)) ##### ODDS RATIO OF THE VARIABLES
OR
```

```
(Intercept)      x
433.0632952    0.9824511
```

Hide

```
#part d
df$xsq=df$x^2
md11b= glm ( y ~ x + xsq, data = df, family = binomial)
summary(md11b) ##### PARAMETER ESTIMATES
```

```
Call:
glm(formula = y ~ x + xsq, family = binomial, data = df)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0642	-0.4879	0.3890	0.5469	2.1651

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.193e+00	9.030e+00	0.686	0.493
x	-1.847e-02	5.513e-02	-0.335	0.738
xsq	1.100e-06	7.921e-05	0.014	0.989

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 34.617 on 24 degrees of freedom
 Residual deviance: 20.363 on 22 degrees of freedom
 AIC: 26.363

Number of Fisher Scoring iterations: 4

Hide

```
#question 1
#part a
library(openxlsx)
xlsxFile <- system.file("extdata","hw_9_data.xlsx", package = "openxlsx")
df <- read.xlsx(xlsxFile = "hw_9_data.xlsx", sheet = 2, skipEmptyRows = FALSE)

mdl2= glm( y ~ x1+x2, data = df, family = binomial)
#part b
summary(mdl2)
```

Call:
glm(formula = y ~ x1 + x2, family = binomial, data = df)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5635	-0.8045	-0.1397	0.9535	1.7915

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.047e+00	4.674e+00	-1.508	0.132
x1	7.382e-05	6.371e-05	1.159	0.247
x2	9.879e-01	5.274e-01	1.873	0.061 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27.726 on 19 degrees of freedom
Residual deviance: 21.082 on 17 degrees of freedom
AIC: 27.082

Number of Fisher Scoring iterations: 5

Hide

```
hoslem.test(df$y,fitted(mdl2), g=8)
```

Hosmer and Lemeshow goodness of fit (GOF) test

data: df\$y, fitted(mdl2)
X-squared = 4.7136, df = 6, p-value = 0.581

Hide

```
#part c
OR=exp(coef(mdl2))
OR
```

```
(Intercept)          x1          x2
0.0008699617 1.0000738195 2.6855513881
```

Hide

```
#part d
nwdt=with(df, data.frame(x1=45000, x2=5))
#####
pct=0.95
#####
nwdt2=subset(cbind(nwdt,predict(md12,newdata=nwdt, type="link", se=TRUE)),select = -c(residual.s
cale))
#####
nwdt3=within(nwdt2,{PredictedProb <- plogis(fit)
  LL <- plogis(fit - (qnorm((1+pct)/2) * se.fit))
  UL <- plogis(fit + (qnorm((1+pct)/2) * se.fit))})
nwdt3
```

	x1 <dbl>	x2 <dbl>	fit <dbl>	se.fit <dbl>	UL <dbl>	LL <dbl>	PredictedProb <dbl>
1	45000	5	1.214124	0.8630815	0.9481291	0.3828464	0.7710279

1 row

Hide

```
#part e
df$x12=df$x1*df$x2
md12b= glm ( y ~ x1 + x2 + x12, data = df, family = binomial)
summary(md12b)
```

Call:

```
glm(formula = y ~ x1 + x2 + x12, family = binomial, data = df)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.63981	-0.62754	-0.05642	0.66213	1.85666

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.144e-01	6.394e+00	0.049	0.961
x1	-1.411e-04	1.412e-04	-0.999	0.318
x2	-2.462e+00	2.081e+00	-1.183	0.237
x12	1.014e-04	6.297e-05	1.610	0.107

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27.726 on 19 degrees of freedom
 Residual deviance: 16.551 on 16 degrees of freedom
 AIC: 24.551

Number of Fisher Scoring iterations: 6

Hide

```
#part f
pct2=0.95
```

```
CI_lower1 = coefficients(md12)[2] - (qnorm((1+pct2)/2))*summary(md12)$coefficients[2,2]
CI_upper1 = coefficients(md12)[2] + (qnorm((1+pct2)/2))*summary(md12)$coefficients[2,2]
```

```
CI_lower1
```

```
x1
-5.105847e-05
```

Hide

```
CI_upper1
```

```
x1
0.0001986921
```

Hide

```
CI_lower2 = coefficients(md12)[3] - (qnorm((1+pct2)/2))*summary(md12)$coefficients[3,3]
CI_upper2 = coefficients(md12)[3] + (qnorm((1+pct2)/2))*summary(md12)$coefficients[3,3]
```

```
CI_lower2
```

```
x2  
-2.683594
```

[Hide](#)

```
CI_upper2
```

```
x2  
4.659366
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.