

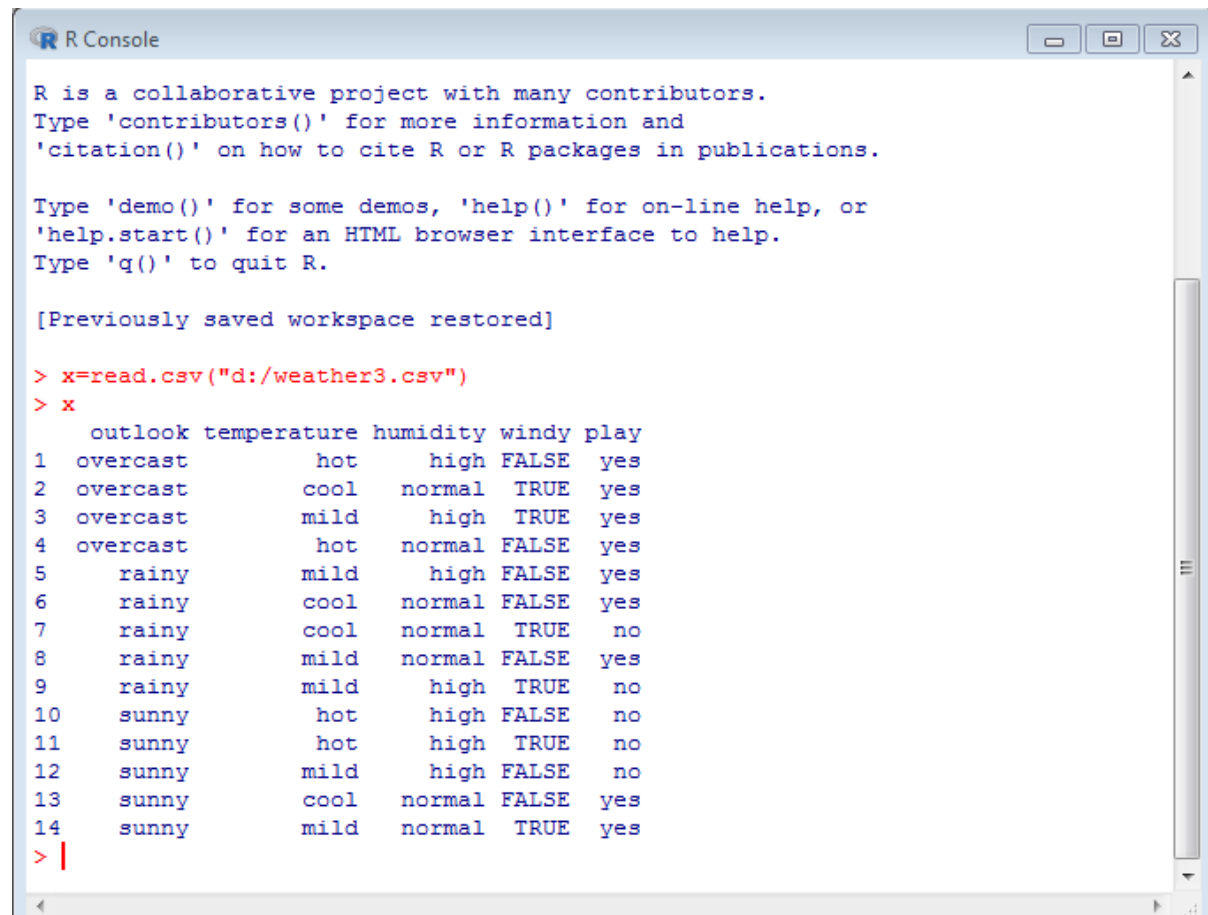
Aim: Demonstration of Logistics Regression.

Theory: Explain Logistic Regression in detail.

Code:

```
X = read.csv("C:/Users/Admin/Documents/SampleStudentData.csv")
```

```
> X
```



```
R Console

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> x=read.csv("d:/weather3.csv")
> x
  outlook temperature humidity windy play
1 overcast         hot      high FALSE  yes
2 overcast         cool    normal  TRUE  yes
3 overcast         mild     high  TRUE  yes
4 overcast         hot     normal FALSE  yes
5  rainy          mild     high FALSE  yes
6  rainy          cool    normal FALSE  yes
7  rainy          cool    normal  TRUE   no
8  rainy          mild    normal FALSE  yes
9  rainy          mild     high  TRUE   no
10 sunny          hot     high FALSE   no
11 sunny          hot     high  TRUE   no
12 sunny          mild     high FALSE   no
13 sunny          cool    normal FALSE  yes
14 sunny          mild    normal  TRUE  yes
> |
```

PRINTING THE DATASET

```
>x$humidity=ifelse(test=x$humidity=="high",yes=1,no=0)
```

```
>x
```

```
> x$humidity=ifelse(test=x$humidity=="high",yes=1,no=0)
```

```
> x
```

```
  outlook temperature humidity windy play
1 overcast         hot        1 FALSE  yes
2 overcast         cool        0  TRUE  yes
3 overcast         mild        1  TRUE  yes
4 overcast         hot         0 FALSE  yes
5  rainy          mild         1 FALSE  yes
6  rainy          cool         0 FALSE  yes
7  rainy          cool         0  TRUE   no
8  rainy          mild         0 FALSE  yes
9  rainy          mild         1  TRUE   no
10 sunny          hot          1 FALSE   no
11 sunny          hot          1  TRUE   no
12 sunny          mild         1 FALSE   no
13 sunny          cool         0 FALSE  yes
14 sunny          mild         0  TRUE  yes
```

```
>x$play=ifelse(test=x$play=="yes",yes=1,no=0)
```

>x

```
> x$play=ifelse(test=x$play=="yes",yes=1,no=0)
> x
```

	outlook	temperature	humidity	windy	play
1	overcast	hot	1	FALSE	1
2	overcast	cool	0	TRUE	1
3	overcast	mild	1	TRUE	1
4	overcast	hot	0	FALSE	1
5	rainy	mild	1	FALSE	1
6	rainy	cool	0	FALSE	1
7	rainy	cool	0	TRUE	0
8	rainy	mild	0	FALSE	1
9	rainy	mild	1	TRUE	0
10	sunny	hot	1	FALSE	0
11	sunny	hot	1	TRUE	0
12	sunny	mild	1	FALSE	0
13	sunny	cool	0	FALSE	1
14	sunny	mild	0	TRUE	1

>x\$windy=ifelse(test=x\$windy=="FALSE",yes=0,no=1)

>x

```
> x$windy=ifelse(test=x$windy=="FALSE",yes=0,no=1)
> x
```

	outlook	temperature	humidity	windy	play
1	overcast	hot	1	0	1
2	overcast	cool	0	1	1
3	overcast	mild	1	1	1
4	overcast	hot	0	0	1
5	rainy	mild	1	0	1
6	rainy	cool	0	0	1
7	rainy	cool	0	1	0
8	rainy	mild	0	0	1
9	rainy	mild	1	1	0
10	sunny	hot	1	0	0
11	sunny	hot	1	1	0
12	sunny	mild	1	0	0
13	sunny	cool	0	0	1
14	sunny	mild	0	1	1

> |

PARTIONING DATASET

```
> s=sample(nrow(x),.7*nrow(x))
```

```
> x_tr=x[s,]
```

```
> x_test=x[-s,]
```

```
> nrow(x)
```

```
> nrow(x_tr)
```

```
> nrow(x_test)
```

```

> s=sample(nrow(x),.7*nrow(x))
> x_tr=x[s,]
> x_test=x[-s,]
> nrow(x)
[1] 14
> nrow(x_tr)
[1] 9
> nrow(x_test)
[1] 5
> |

```

DATA MODELING

```

>lmod=glm(play~windy,data=x_tr,family=binomial,control=list(maxit=100))
>lmod

```

```

> lmod=glm(play~windy,data=x_tr,family=binomial,control=list(maxit=100))
> lmod

```

```

Call:  glm(formula = play ~ windy, family = binomial, data = x_tr, control = list(maxit = 100))

```

Coefficients:

(Intercept)	windy
20.57	-19.87

Degrees of Freedom: 8 Total (i.e. Null); 7 Residual

Null Deviance: 6.279

Residual Deviance: 3.819 AIC: 7.819

```

> |

```

```

> summary(lmod)

```

Call:

```

glm(formula = play ~ windy, family = binomial, data = x_tr, control = list(maxit = 100))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.48230	0.00005	0.00005	0.00005	0.90052

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	20.57	7238.39	0.003	0.998
windy	-19.87	7238.39	-0.003	0.998

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6.2790 on 8 degrees of freedom
 Residual deviance: 3.8191 on 7 degrees of freedom
 AIC: 7.8191

Number of Fisher Scoring iterations: 19

```

> |

```

```

>lmod=glm(play~humidity,data=x_tr,family=binomial,control=list(maxit=100))
>summary(lmod)

```

```
> lmod=glm(play~humidity,data=x_tr,family=binomial,control=list(maxit=100))
> summary(lmod)
```

Call:

```
glm(formula = play ~ humidity, family = binomial, data = x_tr,
     control = list(maxit = 100))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.97277	0.00008	0.55525	0.55525	0.55525

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.792	1.080	1.659	0.0971 .
humidity	17.774	7604.236	0.002	0.9981

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6.2790 on 8 degrees of freedom
 Residual deviance: 5.7416 on 7 degrees of freedom
 AIC: 9.7416

Number of Fisher Scoring iterations: 18

```
> |
```

```
>lmod=glm(play~temperature,data=x_tr,family=binomial,control=list(maxit=100))
>summary(lmod)
```

```
> lmod=glm(play~temperature,data=x_tr,family=binomial,control=list(maxit=100))
> summary(lmod)
```

Call:

```
glm(formula = play ~ temperature, family = binomial, data = x_tr,
     control = list(maxit = 100))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.66511	0.00005	0.00005	0.75853	0.75853

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.099	1.155	0.951	0.341
temperaturehot	19.467	12537.265	0.002	0.999
temperaturemild	19.467	10236.634	0.002	0.998

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6.2790 on 8 degrees of freedom
 Residual deviance: 4.4987 on 6 degrees of freedom
 AIC: 10.499

Number of Fisher Scoring iterations: 19

```
> |
```

#PREDICTION:

```
> p=predict(lmod,x_test,type="response")
>p
```

```
> p=predict(lmod,x_test,type="response")
> p
           3           9           10           11           12
1.000000e+00 5.800756e-11 1.000000e+00 1.000000e+00 1.000000e+00
> |
```

(2) SECOND DATA SET:

#IMPORT THE DATA

```
>x2=read.csv("D:/grade_logit.csv")
```

```
>x2
```

```

> x2=read.csv("D:/grade_logit.csv")
> x2
  Exam1 Exam2 Exam3 Exam4 Final_score Grade
1     60    10    16    7.0     40.79     1
2     90     0     0     0.0     69.23     1
3    130    20    24    1.0     76.75     1
4    130    10    24    8.5     75.66     1
5     90     5    22    9.5     55.48     1
6    100    30    20    3.0     67.11     1
7    105    20    22    8.0     67.98     1
8    120    40    18   16.0     85.09     1
9    120    20    30   18.0     82.46     1
10   130    45    22   10.5     91.01     1
11    90    40    20    7.0     68.86     1
12   130    30    28   10.5     87.06     1
13   100    30    22    6.5     69.52     1
14     0    30    18    0.0     60.00     1
15     0    30    18    0.0     60.00     1
16    80     0    24    3.0     60.11     1
17   105    40    22    6.5     76.10     1
18    10     0     0    8.0     12.16     0
19   130    35    24    0.0     90.00     1
20     0    15    20    7.0     42.86     1
21    40    10    14    6.0     30.70     0
22    90    15    28    8.5     62.06     1
23   110     0    24    9.5     80.62     1
24    65     5    24    1.0     41.67     1
25    55    15    18    0.0     41.90     1
26   100    50    30   11.5     83.99     1
27    95    40    24    8.0     73.25     1
28     0    10    24    0.0     42.50     1
29     0     0    18    0.0     60.00     1
30    65    20    20    0.0     50.00     1
31   110    25    18    6.0     69.74     1
32   130    45    24    8.0     90.79     1
33   120    40    30    9.0     87.28     1
34    70    20    24    1.0     50.44     1
35   130    45    10   16.5     88.38     1
```

```
> lmod2=glm(Grade~Exam1,data=x2_train,family=binomial,control=list(maxit=100))
```

```
>summary(lmod2)
```

```

> lmod2=glm(Grade~Exam1,data=x2_train,family=binomial,control=list(maxit=100))
> summary(lmod2)

Call:
glm(formula = Grade ~ Exam1, family = binomial, data = x2_train,
    control = list(maxit = 100))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2051   0.1834   0.2442   0.4444   0.9351

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.600860   0.396710   1.515  0.12987
Exam1        0.028971   0.009424   3.074  0.00211 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.589  on 82  degrees of freedom
Residual deviance: 54.049  on 81  degrees of freedom
AIC: 58.049

Number of Fisher Scoring iterations: 6

```

Prediction data 1's and 0's form

```
> prediction=ifelse(p>.5,1,0)
```

```
> prediction
```

```

> prediction=ifelse(p>.5,1,0)
> prediction
 4 10 13 14 23 37 45 50 51 55 64 66 67 76 81 84 89 91 93 96 97
1  1  1  1  1  1  1  1  0  1  1  1  0  1  1  1  0  1  1  1  1
> |

```

PREDICTION MATRIX

```
> table(x2_test$Grade,prediction)
```

```

> table(x2_test$Grade,prediction)
      prediction
      0      1
0      2      1
1      1     17

```

```
> x2_test
```

```

> x2_test
  Exam1 Exam2 Exam3 Exam4 Final_score Grade
4    130    10    24    8.5      75.66    1
10   130    45    22   10.5      91.01    1
13   100    30    22    6.5      69.52    1
14     0    30    18    0.0      60.00    1
23   110     0    24    9.5      80.62    1
37     0    25    24    0.0      61.25    1
45    95    30    30   12.0      73.25    1
50   130    40    28   16.5      94.08    1
51     0     0     0   15.5      86.11    1
55   110    25    20    3.0      69.30    1
64   125    30    30   11.5      86.18    1
66    75    15    16    0.0      50.48    1
67     0     0     0    5.0      27.78    0
76   100    35    24    0.0      75.71    1
81    50    20    20    1.0      39.91    0
84   100    35    24   10.5      74.34    1
89     0     0     0    2.0      11.11    0
91   110    25    24    4.0      71.49    1
93    85    30    20    2.5      60.31    1
96   100    35    20    0.0      73.81    1
97     0     0    26    0.0      86.67    1
> |

```

#actuals predicted

```
>ac_pr<- data.frame(cbind(actuals=x2_test$Grade, predicted=prediction))
```

```
>ac_pr
```

```

> ac_pr <- data.frame(cbind(actuals=x2_test$Grade, predicted=prediction))
> ac_pr
  actuals predicteds
4        1          1
10       1          1
13       1          1
14       1          1
23       1          1
37       1          1
45       1          1
50       1          1
51       1          0
55       1          1
64       1          1
66       1          1
67       0          0
76       1          1
81       0          1
84       1          1
89       0          0
91       1          1
93       1          1
96       1          1
97       1          1
> |

```

>vif(lmod2) // variable influence factor

```

> vif(lmod2)
  Exam1    Exam2    Exam3
1.023350 1.117704 1.122152
> |

```