> ##OUR COUNTRY IN FACTS AND FIGURES ##2020##

> ##fitting of logistic regression model

> ##to create the prediction model, we split the data into train -test data(70-30%)

> library(readxl)

> d=read\_excel("C:\\Users\\DELL\\Desktop\\Book1.xlsx")

New names:

\* `` -> `...5`

\* `` -> `...6`

\* `` -> `...7`

\* `` -> `...8`

\* `` -> `...9`

\* `` -> `...10`

> d ##

# A tibble: 27 x 10

state `no.of plants` `co2 emissions` if c02 emissons decr~1

<chr> <dbl> <dbl> <dbl>

1 A& N ~ 142 2.42 1

2 ANDHR~ 233 235. 0

3 ARUNA~ 147 3.3 1

4 ASSAM 186 8.74 1

5 BIHAR 170 5.22 1

6 CHAND~ 112 7.1 1

7 CHHAT~ 193 335. 0

8 DADRA~ 106 5.38 1

9 DELHI 159 10.4 1

10 DIU &~ 108 9.21 1

# i 17 more rows

# i abbreviated name: 1: `if c02 emissons decrease,1 0 o.w.`

# i 6 more variables: ...5 <lgl>, ...6 <dbl>, ...7 <lgl>,

# ...8 <lgl>, ...9 <lgl>, ...10 <dbl>

# i Use `print(n = ...)` to see more rows

> y=d$`if c02 emissons decrease,1 0 o.w.`

> x=d$`no.of plants`

> model=glm(y~x,family="binomial")

Warning message:

glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(model)

Call:

glm(formula = y ~ x, family = "binomial")

Deviance Residuals:

Min 1Q Median 3Q Max

-1.37463 -0.00738 0.00130 0.02994 1.42836

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 40.0363 32.5726 1.229 0.219

x -0.2051 0.1672 -1.226 0.220

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 32.8154 on 26 degrees of freedom

Residual deviance: 7.0335 on 25 degrees of freedom

AIC: 11.033

Number of Fisher Scoring iterations: 10

> s=summary(model)

>

> null=s$coefficients[1]

> null

[1] 40.03635

> residual=s$coefficients[2]

> residual

[1] -0.2050974

> g=null-residual

> g

[1] 40.24145

> tabulated=qchisq(0.95,1)

> tabulated##g>tabulated##regressor is significant.

[1] 3.841459

>

> beta=model$coefficients

> beta

(Intercept) x

40.0363478 -0.2050974

> ##test data

> test=read\_excel("C:\\Users\\DELL\\Desktop\\Book2.xlsx")

New names:

\* `` -> `...5`

\* `` -> `...6`

\* `` -> `...7`

\* `` -> `...8`

> test

# A tibble: 11 x 8

state `no of plants` `co2 emissions` `ifelse command` ...5

<chr> <dbl> <dbl> <dbl> <lgl>

1 LAKSH~ 113 10.4 1 NA

2 MADHY~ 228 134. 0 NA

3 MAHAR~ 267 268. 0 NA

4 MANIP~ 151 3.7 1 NA

5 MEGHA~ 153 9.48 1 NA

6 MIZOR~ 138 4.43 1 NA

7 NAGAL~ 140 5.59 1 NA

8 ODISHA 190 436. 0 NA

9 UTTAR~ 225 101. 0 NA

10 UTTAR~ 191 4.91 1 NA

11 WEST ~ 207 168. 0 NA

# i 3 more variables: ...6 <lgl>, ...7 <lgl>, ...8 <dbl>

> y1=test$`ifelse command`

> x1=test$`no of plants`

> fittedvalues=exp(beta[1]+beta[2]\*x1)/(1+exp(beta[1]+beta[2]\*x1))

> fittedvalues

[1] 1.000000e+00 1.198044e-03 4.028643e-07 9.998846e-01

[5] 9.998260e-01 9.999920e-01 9.999879e-01 7.441852e-01

[9] 2.214363e-03 7.032326e-01 8.174880e-02

> a=ifelse(fittedvalues>0.5,1,0)

> a

[1] 1 0 0 1 1 1 1 1 0 1 0

> y1

[1] 1 0 0 1 1 1 1 0 0 1 0

> accuracy=(length(which(a==y1))/length(a))\*100

> accuracy##the accuracy of the predicted model is 90.9 %

[1] 90.90909

>

> #odds ratio

> estimated\_beta1=beta[2]

> estimated\_beta1

x

-0.2050974

> oddsratio=exp(estimated\_beta1)

> oddsratio##81.4 % chances

x

0.814568

> plot(model)



