**Department of Statistics,**

**Modern college of Arts, Science and Commerce, Pune-05**

**M.Sc.I (Statistics) Semester II**

**ST-25                                                                                                       Date:**

**Practical No. 18                                                                              Submission date:**

**Title : Logistic regression**

Q1.The table below presents the test-firing results for 25 surfaces-to –air anti aircraft missiles at target of varying speed the result of each test is either a hit(y=1) or miss(y=0)

|  |  |  |
| --- | --- | --- |
| Test | target speed (X) in knots | y |
| 1 | 400 | 0 |
| 2 | 220 | 1 |
| 3 | 490 | 0 |
| 4 | 210 | 1 |
| 5 | 500 | 0 |
| 6 | 270 | 0 |
| 7 | 200 | 1 |
| 8 | 470 | 0 |
| 9 | 480 | 0 |
| 10 | 310 | 1 |
| 11 | 240 | 1 |
| 12 | 490 | 0 |
| 13 | 420 | 0 |
| 14 | 330 | 1 |
| 15 | 280 | 1 |
| 16 | 210 | 1 |
| 17 | 300 | 1 |
| 18 | 470 | 1 |
| 19 | 230 | 0 |
| 20 | 430 | 0 |
| 21 | 460 | 0 |
| 22 | 220 | 1 |
| 23 | 250 | 1 |
| 24 | 200 | 1 |
| 25 | 200 | 0 |

1. Fit a logistic regression model to the response variable Y. use a simple linear regression model as the structure for the linear predictor.
2. Does the model deviance indicate that the logistic regression model from part a is adequate?
3. Provide an interpretation of the parameterβ1 in this model.
4. Is there any evidence that quadratic term is required in the model?

Q2 .The compressive strength of an alloy fastener used in aircraft construction is being studied ten loads were selected over the range 2500-4300 psi and a number of fastener failing at each load were recorded. The complete test data are shown below.

|  |  |  |
| --- | --- | --- |
| Load(psi)X | sample size (n) | number failing r |
| 2500 | 50 | 10 |
| 2700 | 70 | 17 |
| 2900 | 100 | 30 |
| 3100 | 60 | 21 |
| 3300 | 40 | 18 |
| 3500 | 85 | 43 |
| 3700 | 90 | 54 |
| 3900 | 50 | 33 |
| 4100 | 80 | 60 |
| 4300 | 65 | 51 |

a. fit a logistic regression model to the data .Use a simple linear regression model as the structure for the linear predictor.

b. Does the model deviance indicate that the logistic regression model from part a is adequate?

c. Expand the predictor to include a quadratic term. Is there any evidence that this quadratic term is required in the model?

d. bFor a quadratic model in part c .Find the wald statistics for each individual model parameter,

Compute 95% confidence interval on the model parameter for quadratic part in c

ALGORITHM

Conditional distribution of response variable is Bernoulli with probability π(x)

Where , π(x) is odds ratio

Therefore the regression model as,

Y=E(Y|X)+ε

= π(x)+ε

Where ε is Bernoulli random variable with E(ε)=0 and V(ε)= π(x) \*(1- π(x))

* Logistic regression model is given as :-

Y= + ε

= π(x)+ε

Where , & are regression coefficients.

* Testing of hypothesis ;-

To test ,

H0: Fitted model is good . v/s

H1:Fitted model is not good .

Model adequacy check by model deviance (λ(β))

λ(β) =2 ln L(Saturated model ) - 2 ln L(Fitted model )

compare λ(β) with χ2α,n-p

* Decision:-

If λ(β) < = χ2α,n-p then we accept H0 i.e. the fitted model is adequate

* Wald’s test:-

For testing ,

H0:βi=0 v/s H1: βi≠0

Test statistics:-

W/ H0= βihat / SE(βihat) ~ Z α/2

Decision:-

If |W| > Z α/2

We reject H0 at 5% of l.o.s

* C.I= (βihat± SE(βihat))

Minitab procedure

1. For dichotomous variable

Stat---regression---Binary logistic regression----fit binary log regression -----select response in a binary response format----in response Y -----continuous predictor X----ok.

1. To include quadratic terms :-

In continuous predictor select X and X^2 ---ok

1. If response is not binary :-

i.ie in success , failure and number of total is give

select:- In binary logistic regression --- select response in event trial format .

in event name success

number of events = n-r

number of trials :- n

repeat this procedure.

**Answers:**

a)Coefficients

Term Coef SE Coef 95% CI Z-Value P-Value VIF

Constant 4.57 1.70 (1.24, 7.90) 2.69 0.007

(X) -0.01364 0.00503 (-0.02350, -0.00377) -2.71 0.007 1.00

Odds Ratios for Continuous Predictors

Odds Ratio 95% CI

target speed (X) in knots 0.9865 (0.9768, 0.9962)

**Regression Equation**

**P(1) = exp(Y')/(1 + exp(Y'))**

**Y' = 4.57 - 0.01364 target speed (X) in knots**

b) Deviance Table

Source DF Seq Dev Contribution Adj Dev Adj Mean Chi-Square

Regression 1 10.47 30.24% 10.47 10.468 10.47

target speed (X) 1 10.47 30.24% 10.47 10.468 10.47

Error 23 24.15 69.76% 24.15 1.050

Total 24 34.62 100.00%

Source P-Value

Regression 0.001

target speed (X) in knots 0.001

Error

Total

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

30.24% **27.35%** 28.15

Goodness-of-Fit Tests

Test DF Chi-Square P-Value

Deviance 23 **24.15** 0.396

Pearson 23 25.47 0.327

Hosmer-Lemeshow 8 6.63 0.577

**Inverse Cumulative Distribution Function .**

Chi-Square with 23 DF

P( X ≤ x ) x

0.95 **35.1725**

**ꭕcal = 24.15 < ꭕ23,0.05 =35.1725 so me accept the null hypothesis and conclude that the model is adequate, but as the adjusted R2 is just 27.35 % which shows that the model is not appropriately fit to the data**.

c) Applying the Wald test to test the significance of the parameter β1.

Here,

β1 = -0.01364 S.E(β1)= 0.00503

W = β1 / S.E(β1)

= -0.01364 / 0.00503

= - 2.711729

**|W| = 2.711729**

**Inverse Cumulative Distribution Function**

Normal with mean = 0 and standard deviation = 1

P( X ≤ x ) x

0.975 **1.95996**

|W| = 2.711729 > Z 0.025 = 1.95996 we reject the null hypothesis and conclude that the target speed (X regressor) is significantly contributing in the model. **For unit increase in the target speed(X) , there is decrease of 0.01364 in the log odds of hitting the target.**

d) Checking for the evidence of the quadratic term in the model :

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

32.37% **26.59%** 29.41

Coefficients

Term Coef SE Coef 95% CI Z-Value P-Value

Constant -1.87 7.70 ( -16.96, 13.21) -0.24 0.808

target speed (X) 0.0281 0.0497 ( -0.0694, 0.1256) 0.57 0.572

X^2 -0.000062 0.000074 (-0.000206, 0.000083) -0.83 0.405

Term VIF

Constant

target speed (X) in knots 88.00

X^2 **88.00**

Odds Ratios for Continuous Predictors

Odds Ratio 95% CI

target speed (X) in knots 1.0285 (0.9330, 1.1338)

X^2 0.9999 (0.9998, 1.0001)

**Regression Equation**

**P(1) = exp(Y')/(1 + exp(Y'))**

**Y' = -1.87 + 0.0281 target speed (X) in knots - 0.000062 X^2**

Goodness-of-Fit Tests

Test DF Chi-Square P-Value

Deviance 22 **23.41** 0.379

Pearson 22 25.87 0.257

Hosmer-Lemeshow 8 10.55 0.229

**Inverse Cumulative Distribution Function**

Chi-Square with 22 DF

P( X ≤ x ) x

0.95 **33.9244**

**ꭕcal = 23.41 < ꭕ22,0.05 =33.9244 so me accept the null hypothesis and conclude that the model is adequate, but as the adjusted R2 is just 26.59 % which decreases more when we add the quadratic term in the model on the other hand VIF increases from 1 to 88 which shows that the model is not adequate.**

Q2)

a)

**Binary Logistic Regression: n-r versus X**

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

99.67% **98.78%** 847.71

Coefficients

Term Coef SE Coef VIF

Constant 5.340 0.546

X -0.001548 0.000158 **1.00**

Odds Ratios for Continuous Predictors

Odds Ratio 95% CI

X 0.9985 (0.9981, 0.9988)

**Regression Equation**

**P(Event) = exp(Y')/(1 + exp(Y'))**

**Y' = 5.340 - 0.001548 X**

b) Deviance Table

Source DF Adj Dev Adj Mean Chi-Square P-Value

Regression 1 112.460 112.460 112.46 0.000

X 1 112.460 112.460 112.46 0.000

Error 8 0.372 0.046

Total 9 112.832

Goodness-of-Fit Tests

Test DF Chi-Square P-Value

Deviance 8 **0.37** 1.000

Pearson 8 0.37 1.000

Hosmer-Lemeshow 5 0.20 0.999

**Inverse Cumulative Distribution Function**

Chi-Square with 8 DF

P( X ≤ x ) x

0.95 **15.5073**

**ꭕcal = 0.37 < ꭕ8,0.05 =15.5073 so me accept the null hypothesis and conclude that the model is adequate**

c) Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

99.75% 97.98% 849.62

Coefficients

Term Coef SE Coef VIF

Constant 4.27 3.64

X -0.00091 0.00217 189.36

X^2 -0.000000 0.000000 189.36

Odds Ratios for Continuous Predictors

Odds Ratio 95% CI

X 0.9991 (0.9949, 1.0033)

X^2 1.0000 (1.0000, 1.0000)

**Regression Equation**

**P(Event) = exp(Y')/(1 + exp(Y'))**

**Y' = 4.27 - 0.00091 X - 0.000000 X^2**

d) Applying the Wald test to test the significance of the parameter β1 and β2 :

Here,

β1 = -0.00091 and S.E(β1)= 0.00217

W = β1 / S.E(β1)

= -0.00091 / 0.00217

= -0.41935

**|W| = 0.41935**

And for, β2 = -0.00091 and S.E(β2)= 0.00217

W = β2 / S.E(β2)

=-0.0000 / 0.0000

= 0

**|W| = 0**

**Inverse Cumulative Distribution Function**

Normal with mean = 0 and standard deviation = 1

P( X ≤ x ) x

0.975 **1.95996**

**For both β1 and β2 ,**

**|W|< Z0.025 which concludes that both the regressors do not contribute significantly in the model.**

Coefficients

Term Coef SE Coef 95% CI Z-Value P-Value VIF

Constant 4.27 3.64 ( -2.87, 11.41) 1.17 0.241

**X** -0.00091 0.00217 **( -0.00515, 0.00334)** -0.42 0.676 189.36

**X^2** -0.000000 0.000000 **(-0.000001, 0.000001)** -0.30 0.766 189.36

Confidence intervals :

C.I (β1) = **( -0.00515, 0.00334)**

C.I(β2) = **(-0.000001, 0.000001)**

**But still we can conclude that the quadratic model is not appropriately fitted to the given data.**