# Department of Statistics,

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

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

**ST-25 Date:**

**Practical No. 8**

# Title : Poisson regression and Generalized linear model

Q1. The table below presents the damage due to the aircraft attack for the 30 attacks Y: Number of locations were damage was inflicted on the aircraft

X1: Types of aircraft(if A-4 type aircraft=0,A-6 type aircraft=1) X2: Bomb load (in tones)

X3: Total months of aircrew experience

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| observation | y | x1 | x2 | x3 |
| 1 | 0 | 0 | 4 | 91.5 |
| 2 | 1 | 0 | 4 | 84 |
| 3 | 0 | 0 | 4 | 76.5 |
| 4 | 0 | 0 | 5 | 69 |
| 5 | 0 | 0 | 5 | 61.5 |
| 6 | 0 | 0 | 5 | 80 |
| 7 | 1 | 0 | 6 | 72.5 |
| 8 | 0 | 0 | 6 | 65 |
| 9 | 0 | 0 | 6 | 57.5 |
| 10 | 2 | 0 | 7 | 50 |
| 11 | 1 | 0 | 7 | 103 |
| 12 | 1 | 0 | 7 | 95.5 |
| 13 | 1 | 0 | 8 | 88 |
| 14 | 1 | 0 | 8 | 80.5 |
| 15 | 2 | 0 | 8 | 73 |
| 16 | 3 | 1 | 7 | 116.1 |
| 17 | 1 | 1 | 7 | 100.6 |
| 18 | 1 | 1 | 7 | 85 |
| 19 | 1 | 1 | 10 | 69.4 |
| 20 | 2 | 1 | 10 | 53.9 |
| 21 | 0 | 1 | 10 | 112.3 |
| 22 | 1 | 1 | 12 | 96.7 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 23 | 1 | 1 | 12 | 81.1 |
| 24 | 2 | 1 | 12 | 65.6 |
| 25 | 5 | 1 | 8 | 50 |
| 26 | 1 | 1 | 8 | 120 |
| 27 | 1 | 1 | 8 | 104.4 |
| 28 | 5 | 1 | 14 | 88.9 |
| 29 | 5 | 1 | 14 | 73.7 |
| 30 | 7 | 1 | 14 | 57.8 |

1. Fit a Poisson regression model to the data.
2. Does the model deviance indicate that the Poisson regression model from part a is adequate?
3. Construct the graph of the fitted model versus months. Also plot the observed number of failure on this graph.
4. Expand the linear predictor to include a quadratic term. Is there any evidence that this quadratic term is required in the model?
5. For the model in part a , find Wald statistics for each individual model parameter.

Q2 .A chemical manufacturer has maintained records on the number of failure of particular types

of valve used in its processing unit and the length of time (months) since the valve was installed . The data are shown below.

|  |  |  |
| --- | --- | --- |
| valve | No. of failure | months |
| 1 | 5 | 18 |
| 2 | 3 | 15 |
| 3 | 0 | 11 |
| 4 | 1 | 14 |
| 5 | 4 | 23 |
| 6 | 0 | 10 |
| 7 | 0 | 5 |
| 8 | 1 | 8 |
| 9 | 0 | 7 |
| 10 | 0 | 12 |
| 11 | 0 | 3 |
| 12 | 1 | 7 |
| 13 | 0 | 2 |
| 14 | 7 | 30 |
| 15 | 0 | 9 |

1. Fit a Poisson regression model to the data.
2. Does the model deviance indicate that the Poisson regression model from part a is adequate?
3. Construct the graph of the fitted model versus months. Also plot the observed number of failure on this graph.
4. Expand the linear predictor to include a quadratic term. Is there any evidence that this quadratic term is required in the model?
5. For the model in part a, find Wald statistics for each individual model parameter.

**ALGORITHM**

* If response variable of interest is not normally distributed then the order of regression model is is used to call “generalized linear model”.
* The situation where response variable represents count of relatively rare events such as defective unites in manufacturing product we assume that response variable yi is count such that the observation yi =0,1,2,3…
* A responsible probability model for count dada is often poisson distribution.
* The link function for poisson distribution is log-link function that is ηi= lnλi=λi`β
* Poisson regression model is

yi = exp{ β0 + β1 X1 + β2 X2+ β3 X2+…+ βkXk} or

yi =e^( X1 ` β )

* To check the significance of quadratic terms add quadratic terms ( i.e. X12 , X22…) in the mode and fit the model.
* **Testing of hypothesis:-**

To test -

H0= Fitted model is adequate . v/s

H1 = Fitted model is not adequate .

Model adequacy is check by model deviance i.e.

λ(β) =2\*ln l(standard model ) - 2\*ln L(βhat)

compare λ(β) with χ2 (n-p),α

* **Decision:-**

If λ(β)< χ2table

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

.

* **To check the significance of individual by Wald test :-**

To test -

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

**Solution**

**Q1)**

**a)**

**Poisson Regression Analysis: y versus x1, x2, x3**

Coefficients

Term Coef SE Coef VIF

Constant -0.406 0.877

x1 0.569 0.504 1.99

x2 0.1654 0.0675 2.02

x3 -0.01352 0.00828 1.05

**Regression Equation**

y = exp(Y')

Y' = -0.406 + 0.569 x1 + 0.1654 x2 - 0.01352 x3

**Y= exp {β0 + β1 x1+ β2 x2+ β3 x3}**

Y: Number of locations were damage was inflicted on the aircraft

X1: Types of aircraft (if A -4 type aircraft=0, A-6 type aircraft=1)

X2: Bomb load (in tons)

X3: Total months of aircrew experience

Fits and Diagnostics for Unusual Observations

Std

Obs y Fit Resid Resid

16 3.000 0.779 1.909 2.06 R

R Large residual

**b) Testing of hypothesis:-**

To test -

H0= Fitted model is adequate . v/s

H1 = Fitted model is not adequate .

Deviance Table

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

Regression 3 27.930 9.3100 27.93 0.000

x1 1 1.267 1.2667 1.27 0.260

x2 1 6.239 6.2386 6.24 0.012

x3 1 2.681 2.6811 2.68 0.102

Error 26 25.953 0.9982

Total 29 53.883

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

51.83% 46.27% 87.65

Goodness-of-Fit Tests

Test DF Estimate Mean Chi-Square P-Value

Deviance 26 25.95316 0.99820 25.95 0.466

Pearson 26 23.93838 0.92071 23.94 0.579

**Inverse Cumulative Distribution Function**

Chi-Square with 26 DF

P( X ≤ x ) x

0.95 38.8851

**Decision:-**

If λ(β)< χ2table

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

χ2**cal =**25.95 **<** χ2**26,0.05 =**38.8851**so me accept the null hypothesis and conclude that the model is adequate, but as the adjusted R2 is just** 46.27% **which shows that the model is not appropriately fit to the data**.

**c)**

****

**Here, we can observe that as the damage locations inflicted on the aircraft decreases there is increase in the months of experience or life of aircrew this implies there is near linear relationship in the variables but in opposite direction.**

d) Fitting of the Quadratic terms to see whether there is any evidence for the model to become adequate :

**Poisson Regression Analysis: Y versus X1, X2, X3, X2^2, X3^2**

Method

Link function Natural log

Rows used 30

Deviance Table

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

Regression 5 31.2869 6.25738 31.29 0.000

X1 1 0.1689 0.16894 0.17 0.681

X2 1 0.0064 0.00639 0.01 0.936

X3 1 3.8983 3.89825 3.90 0.048

X2^2 1 0.3528 0.35277 0.35 0.553

X3^2 1 3.2403 3.24034 3.24 0.072

Error 24 22.5961 0.94151

Total 29 53.8831

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

58.06% 48.79% 88.29

Coefficients

Term Coef SE Coef VIF

Constant 5.23 3.47

X1 0.258 0.631 3.11

X2 -0.036 0.455 94.79

X3 -0.1416 0.0707 96.12

X2^2 0.0128 0.0214 85.78

X3^2 0.000801 0.000438 100.80

**Regression Equation**

**Y = exp(Y')**

**Y' = 5.23 + 0.258 X1 - 0.036 X2 - 0.1416 X3 + 0.0128 X2^2 + 0.000801 X3^2**

Checking the adequacy of the model using the deviance test ,

**Hypothesis**

H0: Fitted model is good after adding quadratic terms

V/s H1: Fitted model is not good after adding quadratic terms

**Test Statistics**

D = 2\*ln\* L (Saturated Model) – 2\*ln L (β)

D = 22.60

**Test Procedure**

We reject H0 at 5% l.o.s. if |D| > Ϫ2n-p, α otherwise we accept it.

Goodness-of-Fit Tests

Test DF Estimate Mean Chi-Square P-Value

**Deviance 24** 22.59613 0.94151 **22.60** 0.544

Pearson 24 18.76728 0.78197 18.77 0.764

**Inverse Cumulative Distribution Function**

Chi-Square with 24 DF

P( X ≤ x ) x

0.95 **36.4150**

Here |D|=22.60 < 36.4150, we accept H0.

**Conclusion:**Fitted model is good after adding quadratic terms that is our model is adequate.

**After adding quadratic terms to the model:**

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

58.06% **48.79%** 88.29

**Before adding the quadratic term:**

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

51.83% **46.27%** 87.65

**Here, we can observe that even if Adj R2 has increased we can say that the adequacy of the model has increased but the model is surely not a good fit to the given data.**

e) **To check the significance of individual by Wald test :-**

To test -

H0:- βi =0 v/s

H1 :- βi≠0

Here,

β1 = 0.569 S.E(β1)= 0.504

W = β1 / S.E(β1)

=0.569/0.504

**|W|** =1.12896

**Inverse Cumulative Distribution Function**

Normal with mean = 0 and standard deviation = 1

P( X ≤ x ) x

0.975 **1.95996**

|W| = 1.12896<Z 0.025 = 1.95996 we accept the null hypothesis and conclude that the Types of aircraft(if A-4 type aircraft=0,A-6 type aircraft=1 (X1 regressor) is **not** significantly contributing in the model.

Here,

β2 = 0.1654 S.E(β2)= 0.0675

W = β1 / S.E(β1)

=0.1654/0.0675

**|W|** =2.45037

**Inverse Cumulative Distribution Function**

Normal with mean = 0 and standard deviation = 1

P( X ≤ x ) x

0.975 **1.95996**

|W| = 2.45037 >Z 0.025 = 1.95996 we reject the null hypothesis and conclude that the Bomb load (in tones) (X2regressor) is significantly contributing in the model.

Here,

β2 = -0.01352 S.E(β2)= 0.00828

W = β1 / S.E(β1)

=-0.01352 / 0.00828

**|W|** =1.63285

**Inverse Cumulative Distribution Function**

Normal with mean = 0 and standard deviation = 1

P( X ≤ x ) x

0.975 **1.95996**

|W| = 1.63285 >Z 0.025 = 1.95996 we accept the null hypothesis and conclude that Total months of aircrew (X3 regressor) is **not** significantly contributing in the model.

**Q2)**

**a) Poisson Regression Analysis: No. of failure versus months**

Coefficients

Term Coef SE Coef VIF

Constant -1.720 0.558

months 0.1306 0.0243 1.00

**Regression Equation**

No. of failure = exp(Y')

Y' = -1.720 + 0.1306 months

Y= exp {β0 + β1 x1}

Y= exp {-1.720+ 0.1306 x1}

**b) Testing of hypothesis:-**

To test -

H0= Fitted model is adequate . v/s

H1 = Fitted model is not adequate

Deviance Table

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

Regression 1 29.23 29.233 29.23 0.000

months 1 29.23 29.233 29.23 0.000

Error 13 14.93 1.149

Total 14 44.17

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

66.19% 63.92% 38.48

Goodness-of-Fit Tests

Test DF Estimate Mean Chi-Square P-Value

Deviance 13 14.93496 1.14884 14.93 0.311

Pearson 13 13.33120 1.02548 13.33 0.423

Fits and Diagnostics for Unusual Observations

No. of

Obs failure Fit Resid Std Resid

14 7.00 9.02 -0.70 -1.67 X

X Unusual X

**Inverse Cumulative Distribution Function**

Chi-Square with 13 DF

P( X ≤ x ) x

0.95 22.3620

**Decision:-**

If λ(β)< χ2table

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

χ2**cal =** 14.93**<** χ2**13,0.05 =**22.3620**so me accept the null hypothesis and conclude that the model is adequate, but as the adjusted R2 is just** 63.92% **which shows that the model is not appropriately fit to the data**.

c)



**Here it can be observed that increase in the months results in an increase in the failure rate ,but there is non-linear on curvilinear relationship between the two variables which demands for the quadratic Poisson regression model.**

d) Fitting of the Quadratic terms to see whether there is any evidence for the model to become adequate :

**Poisson Regression Analysis: No. of failure(Y) versus Months(X), X^2**

Method

Link function Natural log

Rows used 15

Deviance Table

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

Regression 2 33.398 16.6990 33.40 0.000

Months(X) 1 8.603 8.6029 8.60 0.003

X^2 1 4.166 4.1655 4.17 0.041

Error 12 10.769 0.8975

Total 14 44.167

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

75.62% 71.09% 36.32

Coefficients

Term Coef SE Coef VIF

Constant -4.44 1.71

Months(X) 0.459 0.180 37.06

X^2 -0.00826 0.00435 37.06

**Regression Equation**

**No. of failure(Y) = exp(Y')**

**Y' = -4.44 + 0.459 Months(X) - 0.00826 X^2**

Goodness-of-Fit Tests

Test DF Estimate Mean Chi-Square P-Value

**Deviance12**  10.76943 0.89745 **10.77** 0.549

Pearson 12 10.62424 0.88535 10.62 0.561

**Inverse Cumulative Distribution Function**

Chi-Square with 12 DF

P( X ≤ x ) x

0.95 **21.0261**

Here |D|=10.77<21.0261, we accept H0.

**Conclusion:**Fitted model is good after adding quadratic terms that is our model is adequate.

**After adding quadratic terms to the model:**

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

75.62% **71.09%** 36.32

**Before adding the quadratic term:**

Model Summary

Deviance Deviance

R-Sq R-Sq(adj) AIC

66.19% **63.92%** 38.48

**Here, we can observe that the adj. R2  has increased from 63.92% to 71.09% which implies that this model is considerably good fitted to the given data and the adequacy increased as we add the quadratic term to the model.**

e) **To check the significance of individual by Wald test :-**

To test -

H0:- βi =0 v/s

H1 :- βi≠0

Here,

β1 = 0.1306 S.E(β1)= 0.0243

W = β1 / S.E(β1)

= 0.1306 / 0.0243

**|W|** =5.37448

**Inverse Cumulative Distribution Function**

Normal with mean = 0 and standard deviation = 1

P( X ≤ x ) x

0.975 **1.95996**

|W| =5.37448 >Z 0.025 = 1.95996 we reject the null hypothesis and conclude that the length of time (months) (X1 regressor) is significantly contributing in the model.