**Date:** 29 October, 2015

**Experiment No. 12**

**Aim:** To obtain predicted response values and test the significance of model parameter for logistic model.

**Experiment:**  The given data is for the number of deaths among the patients with specific scores on a diagnostic test along with the number of patients to which the test was determined:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Scores | No. of patients | No. of deaths | Test Scores | No. of patients | No. of deaths |
| 0 | 1 | 0 | 20 | 13 | 6 |
| 2 | 1 | 0 | 21 | 17 | 9 |
| 3 | 4 | 1 | 22 | 14 | 12 |
| 4 | 11 | 0 | 23 | 13 | 7 |
| 5 | 9 | 3 | 24 | 11 | 8 |
| 6 | 14 | 3 | 25 | 12 | 8 |
| 7 | 12 | 4 | 26 | 6 | 2 |
| 8 | 22 | 5 | 27 | 7 | 5 |
| 9 | 33 | 3 | 28 | 3 | 1 |
| 10 | 19 | 6 | 29 | 7 | 4 |
| 11 | 31 | 5 | 30 | 5 | 4 |
| 12 | 17 | 5 | 31 | 3 | 3 |
| 13 | 32 | 13 | 32 | 3 | 3 |
| 14 | 25 | 7 | 33 | 1 | 1 |
| 15 | 18 | 7 | 34 | 1 | 1 |
| 16 | 24 | 8 | 35 | 1 | 1 |
| 17 | 27 | 8 | 36 | 1 | 1 |
| 18 | 19 | 13 | 37 | 1 | 1 |
| 19 | 15 | 7 | 38 | 1 | 0 |

A logistic regression model is fitted to the data for death of a patient and the parameters was estimated to be = -2.2903 and = 0.1156.

Obtain the fitted number of deaths and test whether

H0: β1 = 0 against H1: β1 ≠ 0. Interpret you results.

**Theory:**

The idea to solve this problem is to replace 𝑿𝒊T𝜷 by some other function of it in the model 𝑌𝑖 = 𝑿𝒊𝑻𝜷 +𝜖𝑖 which can represent the conditional expectation or equivalently conditional probability of success of 𝑌𝑖, i.e.

𝑌𝑖 = (𝑌 𝑖|𝑋1,…,) + 𝜖𝑖

⇔ 𝑖 = (𝑌 𝑖 = 1|𝑋1,…,) + 𝜖𝑖

Some function of 𝑿𝒊𝑻𝜷 which is bound to lie between 0 and 1 will serve the purpose. Generally when response variable is binary, there is considerable empirical evidence indicating that the relationship between (𝑌𝑖 = 1|𝑋1,…,) and 𝑋1,…,𝑋𝑝 is non-linear.

Hence we model the conditional probabilities using a non-linear function of the independent variables of the following form. (𝑌𝑖 = 1|𝑋1,…,) = 𝑔(𝑿𝒊𝑻𝜷)

The function (.) is called as a link function. The most common link function is the logit link function or logistic function or sigmoid function which is defined as follows.

𝑔(𝑧) =

In terms of the logit link function the model under study is actually given by :

𝑌𝑖 = + 𝜖𝑖

Observe that the above expression specifies a non-linear regression model which occurs when the dependent variable depends on independent variables through a non-linear function of unknown parameters. The stochastic model defined is called as a logistic regression model.

To test the null hypothesis we use the test statistic:

z = ()2

which under the null hypothesis follows a chi-square distribution with 1 degrees of freedom.

()2 = (2, 2)th element of (X’VX)-1, and S=( X’VX).

V is the matrix whose diagonal elements are the variances of observed y values.

Reject the null hypothesis if the value of the test statistic is greater than χ20.025;1 or less than χ20.975;1 where χ20.025;1 and χ20.975;1 are the upper 0.025 and lower 0.025 of chi-square distribution with 1 degrees of freedom.

**Algorithm:**

1. Open the file “in12.txt” to read the data and “out12.txt” to write the results using pointers.
2. We formulated inverse() function to compute matrix inverse.
3. We input the given values of response variable, ni’s and the regressor variable from the input file.
4. Then we predicted the number of deaths using the logit link (Theory).
5. Then compute the S matrix, its inverse and the test statistic.
6. Results are expected in the file “out12.txt”.

**Additional:**

χ20.025; 1 = 5.023886 [excel command: chiinv(0.025,1)]

χ20.975; 1 = 0.000982 [excel command: chiinv(0.975,1)]

**Results:**

The value of the test statistic comes out to be 51.960030 and the predicted numbers of deaths are shown in the output file attached alongside.

**Conclusion:**

The value of the test statistic is greater than the upper critical value and hence, the null hypothesis is rejected. Thus, β1 is significant.