Problem set 4

ON FITTING OF GLM

Question 1: You are provided with data on literacy status of n individuals of a village. If a person is literate we designate the value 1 otherwise 2. Along with literacy status you are provided with information on individuals gender (1=male, 2=female) and age. Rename 2 as zero. Delete missing observations if any.

(i) Using gender as the covariate form a 2x2 contingency table. Find the (a) odds of a male to be literate (b) odds of a female to be literate (c) odds ratio. Comment on your findings.

(ii) Fit a logistic regression model using age as the covariate.

(iii) Fit a probit regression model using age as the covariate

(iv) Compute the goodness of fit of model (ii) and (iii). Compare your results and comment.

(v) For model in (ii) and (iii) plot the fitted probabilities versus the values of the predictor (Display in separate panel of the same graph) .

(vi) Using different threshold values obtain the predicted value of Y and hence create the confusion matrix using Y and $\hat{Y}$. Draw the ROC. (Do this for any one model of your choice from (ii)-(iii))

Question 2: Perform the following simulation study and interpret your findings. (Set seed as 987654321)

Step 1:. Generate $x_1, x_2, .., x_n$ from Uniform [0,1]. Compute the predictor $\eta(x) =$

$\beta_0 + \beta x$. Take $\beta_0 = -2, \beta = 1.2$. Using logit link function generate $Y_1, Y_2, .., Y_n$

Step 2: Fit a logistic regression model by solving the score equations numerically.

Step 3: Compute the maximum likelihood estimates of the parameters and the standard error of the estimators by inverting Fisher Information matrix. Also compute the 95% confidence interval for β and the empirical coverage. Further compute the simulated standard error, empirical bias and mean square error.

Repeat steps 1-3, R=1000 times.

For $n = 100, 500, 1000$ report the following in a neat table.

MLE of β, standard error of $\hat{\beta}$ (both analytical and simulated), confidence interval of β, empirical coverage, simulated bias and MSE of $\hat{\beta}$.