1.General Overview

The data from Table 1'credit-score' describes the consumer behaviour and application status(credit acceptance) of 100 consumers with the help of 6 variables: Age, Inc, Home, SelfEm, Der, Appln. Variable 'Age' refers to Consumer age, Variable 'Inc' refers to dependent income, Variable 'Home' describes whether they have a home or not, Variable 'SelfEm' refers to whether the consumer is a self-employee or not, Variable 'Der' refers to the number of insolvencies of consumer within 60 or more days in a credit account, Variable 'Appln' refers to the status of application whether it is accepted or rejected. Age, Inc, Der are continuous variables and Home, SelfEm, Appln are categorical variables.

Here Variable 'Appln' is the response variable (binary dependent variable) which gives only two possible outcomes (1- Accepted, 0 – Rejected). So, we use Logistic Regression to predict the probability of application status (Accepted/Rejected) depending on the variables and find variables which have the strongest relationship to credit acceptance. We will use Poisson Regression, when the dependent variable outcomes are counts.

2.Forward Selection

Model $M0$ is the null model. Model $M1$ used to predict the application status($Appln$) with variable Age.Model $M2$ predicts the application status with variable$Income$. Model $M3$ predicts the application status with variable $Home$. Model $M4$ predicts the application status with variable $SelfEm$. Model $M5$ predicts the application status with variable Der.

Model	G^2	**Degrees of freedom**
$M0$	120.43	99
$M1$	119.32	98
$M2$	114.40	98
$M3$	119.99	98
$M4$	115.77	98
$M5$	108.38	98
$M15$	107.30	97
$M25$	93.667	97
$M35$	107.15	97
$M45$	104.11	97

$M125$	89.823	96
$M235$	93.66	96
$M245$	87.175	96

We compute conditional statistics:

$$G^2(M0) - G^2(M1) = 1.11;\ d.f = 1; p - value = 0.292082$$

$$G^2(M0) - G^2(M2) = 6.03;\ d.f = 1; p - value = 0.014065 *$$

$$G^2(M0) - G^2(M3) = 0.44;\ d.f = 1; p - value = 0.507122$$

$$G^2(M0) - G^2(M4) = 4.66;\ d.f = 1; p - value = 0.030873 *$$

$$G^2(M0) - G^2(M5) = 12.05;\ d.f = 1; p - value = 0.000518 *$$

Model $M2, M4, M5$ improves the fit, and among them $M5$ gives the best improvement.

$$G^2(M5) - G^2(M15) = 1.08;\ d.f = 1; p - value = 0.298698$$

$$G^2(M5) - G^2(M25) = 14.713;\ d.f = 1; p - value = 0.000125 *$$

$$G^2(M5) - G^2(M35) = 1.23;\ d.f = 1; p - value = 0.267407$$

$$G^2(M5) - G^2(M45) = 4.27;\ d.f = 1; p - value = 0.038791 *$$

Model $M25, M45$ improves the fit, and Model $M25$ gives the best improvement.

$$G^2(M25) - G^2(M125) = 3.844;\ d.f = 1; p - value = 0.049924 *$$

$$G^2(M25) - G^2(M235) = 0.007;\ d.f = 1; p - value = 0.933322$$

$$G^2(M25) - G^2(M245) = 6.492;\ d.f = 1; p - value = 0.010836 *$$

Model $M125, M245$ improve the fit. Hence, we choose Model $M245$ as the best model for further calculation.

3. Calculations and Predictions

We have selected the model $M245$ with variables income, self-employment, Number of notices of insolvency to predict the status of credit acceptance.

Fitted curve, $p(x) = \frac{e^{-0.6886+0.8278i-1.7575s-1.2533d}}{1+e^{-0.6886+0.8278i-1.7575s-1.2533d}}$

Therefore, we have:

$$Logit(Y = 1|i, s, d) = \alpha + \beta_1\, i + \beta_{2s} + \beta_3 d\ , s = 1,2, \beta_{21} = 0$$

From the R-output, the coefficient of Der is negative, which indicates a reduction will give a chance of credit acceptance.

Comparing the odds of Der for a customer with 3 more than other customer (holding other variables as fixed).

$log\ \frac{\varphi(i,s,d+3)}{\varphi(i,s,d)} = \beta_3(d+3) - \beta_3 d. Therefore, 3\hat{\beta}_3 = -3.7599$(estimate from R listing).

That is odd of Der for a customer with Der of 3 more than another customer decreased by a factor of -3.7599.

Hence using R, $e^{3\widehat{\beta}_3} = 0.0233$. Corresponding 95% confidence interval is therefore,

$$e^{3(\widehat{\beta}_3 \pm 1.96 * se(\widehat{\beta}_3))} = e^{3((-1.2533)\ \ \pm 1.96*(0.3703))} = (0.2054{,}0.00264)$$

It doesn't contains 1.

Comparing the odds of $Income$ for a customer with 2 less than other customer (holding other variables as fixed).

$log\ \frac{\varphi(i-2,s,d)}{\varphi(i,s,d)} = \beta_1(i-2) - \beta_1 i. Therefore, -2\hat{\beta}_i = -1.6556.$(estimate from R listing).

That is odd of $Income$ for a customer with $Income$ of 2 less than another customer decreased by a factor of -1.6556.

Hence using R, $e^{-2\widehat{\beta}_1} = 0.1909$. Corresponding 95% confidence interval is therefore,

$$e^{-2(\widehat{\beta}_1 \pm 1.96 * se(\widehat{\beta}_1))} = e^{-2((0.8278)\ \ \pm 1.96*(0.2698))} = (0.066323{,}0.5499)$$

It doesn't contains 1.