**MOM-PBD-3803**

Date: 17/08/2021

The data “German Credit Rating” gives information about 800 applicant’s Credit Classification (Credit Rating) along with 8 attributes. The response variable Y takes value 1(Bad Credit) or 0 (Good Credit). The variable descriptions are as follows.

* Duration : Duration of the credit given in months
* Credit Amount : Amount of credit/loan
* Balance in Savings A/C
* Install\_rate : Installment rate
* Present Resident : Residing since in years
* Age : Age in years
* Other installment: Other installment (yes =1 and no=0)
* Num\_Credits : Number of existing credits at this bank
* Credit classification : Good/Bad
* Credit Rating : 1(Bad Credit) or 0 (Good Credit).

Attempt the following questions as part of model building.

**a)** Obtain proportion of applicants having good credit rating and bad credit rating based on the data given?

**b)** Obtain a table for counts of Credit Rating and Other Instalments?

**c)** Do basic visualization on the variable Credit Rating based on other variables given in the data (which you feel suitable to study) and comment your findings.

**d)** Check multicollinearity using VIF and remove those predictors (one at a time) during the procedure if it exists.

Divide the data into training and testing. Attempt the following questions using training data.

**e)** Perform a logistic Regression model to predict credit scoring based on the variable “Other Instalments”. Do you feel that this variable can be considered as a significant predictor for credit rating? Justify.

**f)** Consider the polychotomous categorical predictor “Balance in Savings Account” which has 5 levels namely " less100DM"," over1000DM"," unknown" ,"Between 100 and 500 DM" and "Between 500 and 1000 DM".

Combine the three levels of the variable as follows.

“less100DM”,“Between 100 to 500 DM” and “Between 500 to 1000 DM” together as

“Below 1000 DM”

Perform a logistic regression model to predict credit scoring based on the variable “Balance in Savings Account” which has now 3 levels namely “over1000DM"," unknown", "Below 1000 DM". Does the variable “Balance in Savings Account” is significant for predicting credit rating? Why/Why not? If the variable is significant, write down the estimated model.

**g)** Build a logistic regression model with multiple predictors “Duration, Credit Amount, Instalment Rate, Present Resident, Age and Number of Credits ”.Check the overall model significance. Remove any insignificant variable (one at a time) during the procedure of building the model. Report the predictors which are removed while building the model with proper justification of removal. Write the logit and estimated model.

**h)** What is the probability of credit score of a random applicant if he has 32 years of age, 36 months of credit duration, Rs.2500 credit amount and an instalment rate 2? What would be your suggestion about the credit score based on this probability?

**i)** Comment on the accuracy of the fitted model?

**J)** Prepare a report based on the findings above.