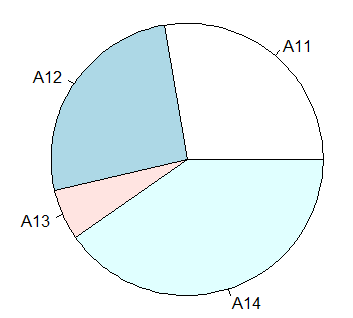
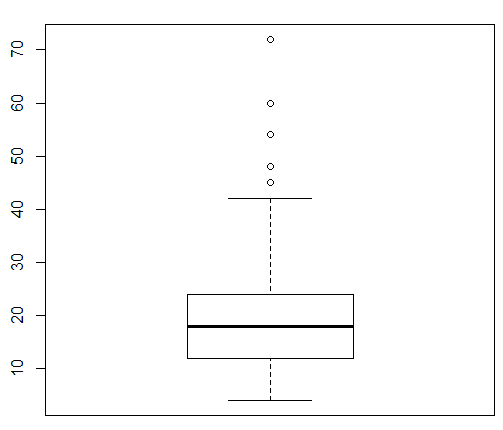
Case 3: Credit Scoring Data

1. Exploratory Analysis of 80% of Original Data
   1. Visual Representation

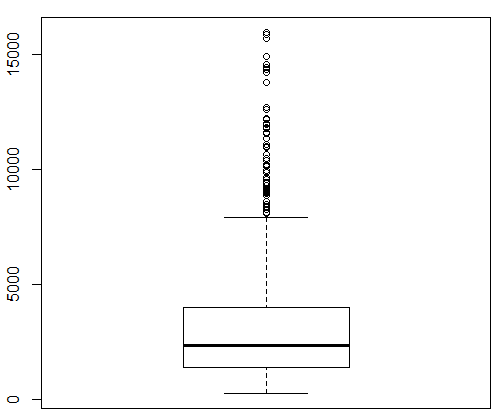
Checking account



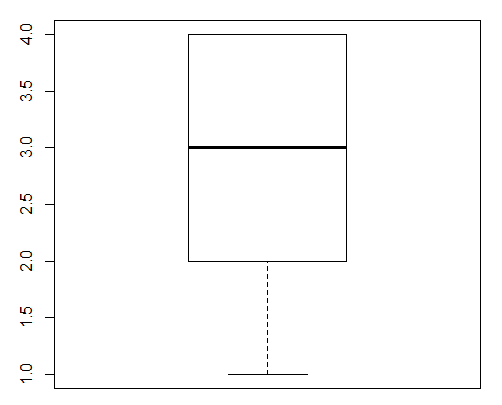
Duration



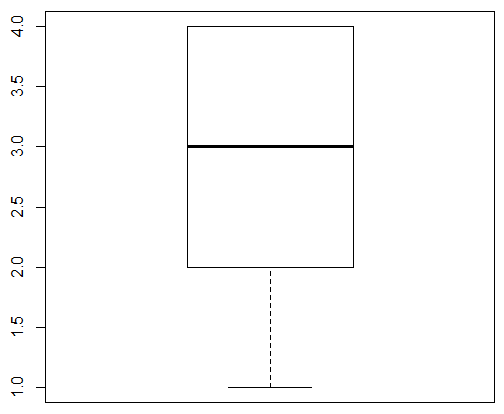
Amount



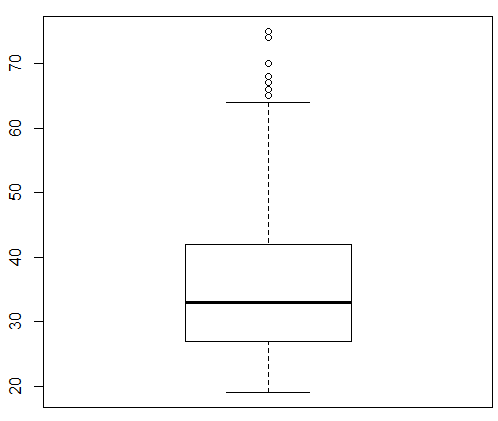
Installment rate



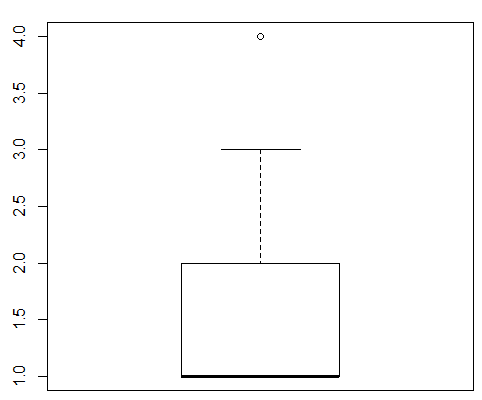
Present\_resid



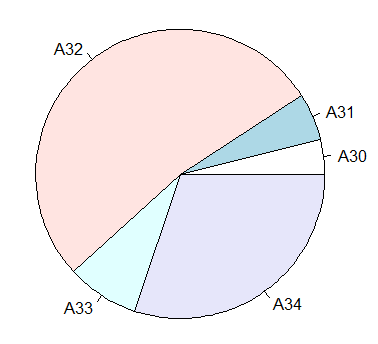
Age



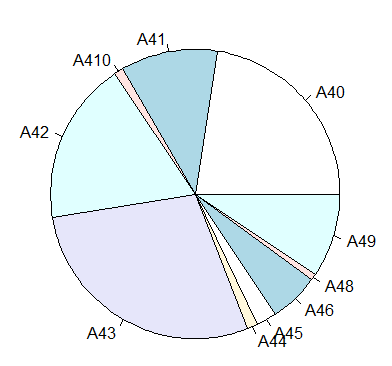
N\_credit



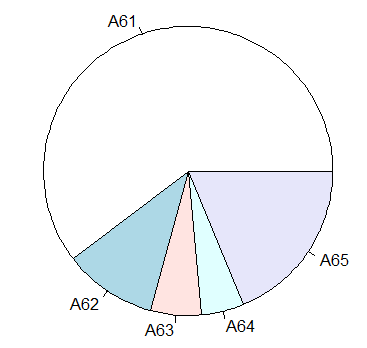
Credit\_his



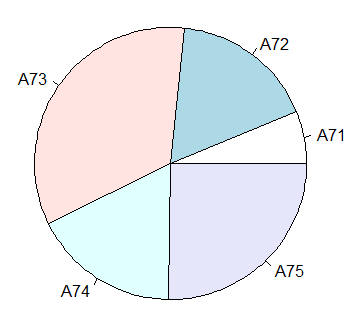
Purpose



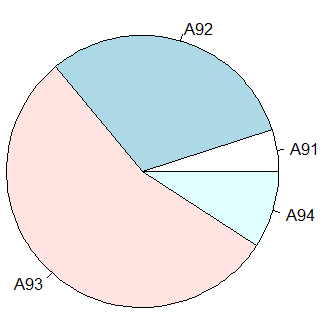
Saving account



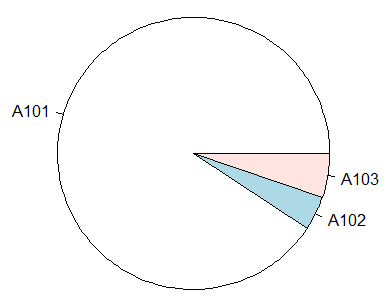
Present\_emp



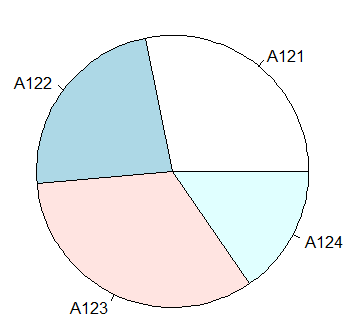
Sex



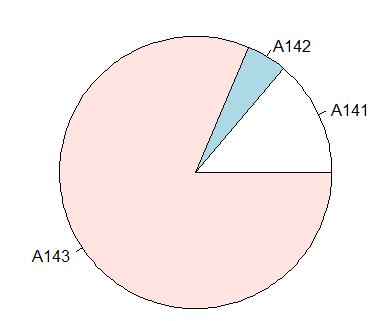
Other\_debtor



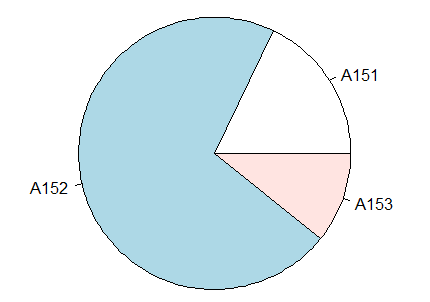
Property



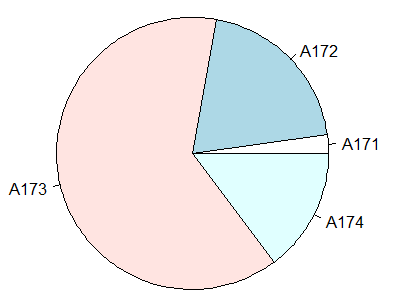
Other\_install



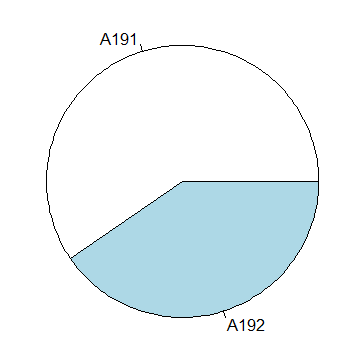
Housing



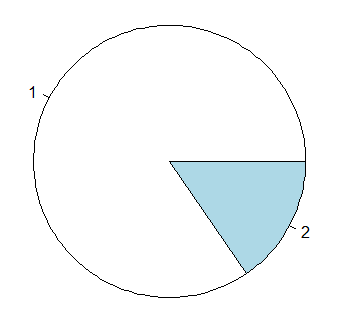
Job



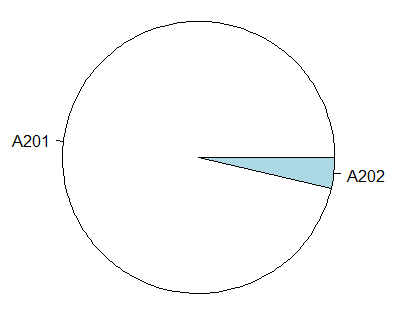
Telephone



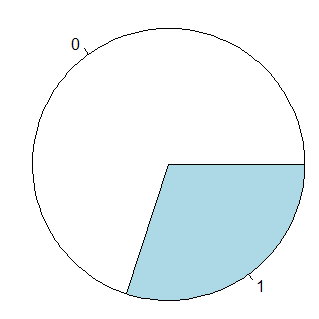
n\_people



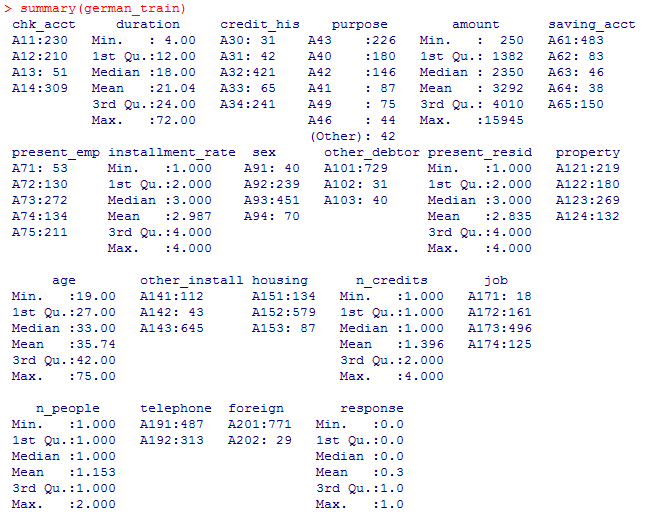
Foreign



Response

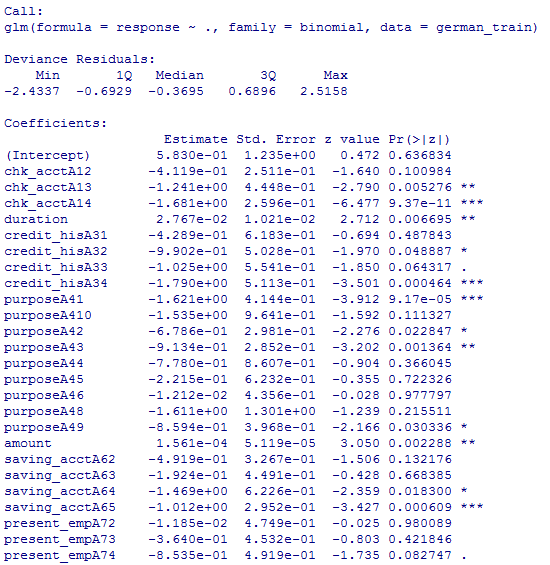


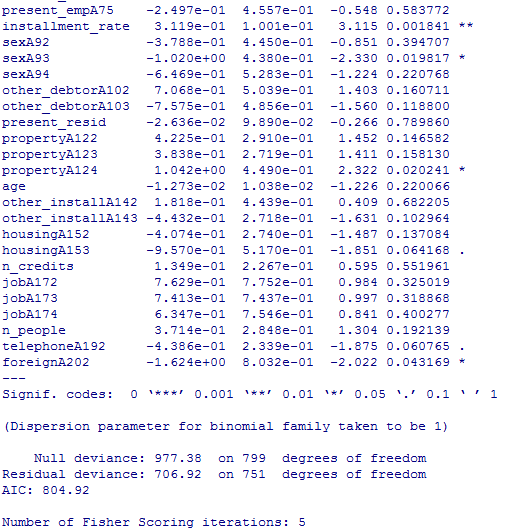
Summary



Logistic regression model

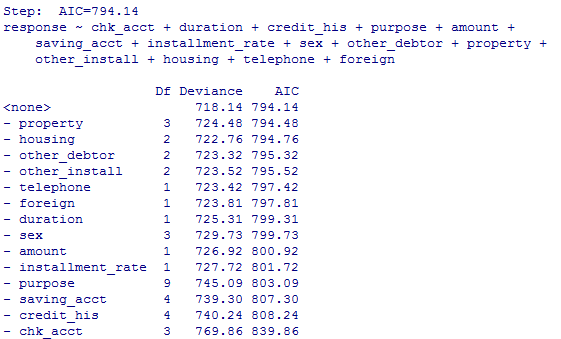






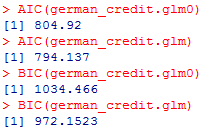
Step wise



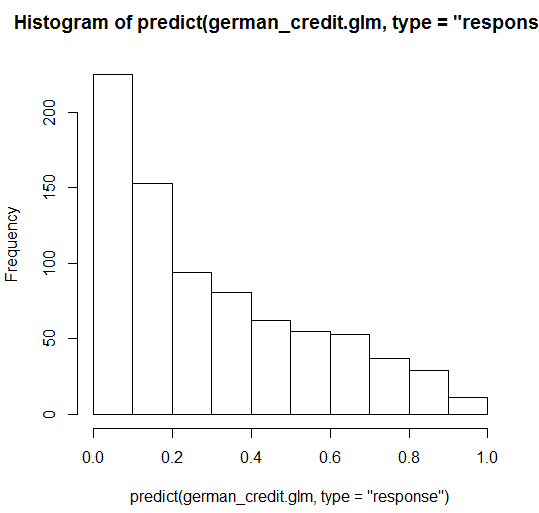


German\_credit.glm0 has all variables

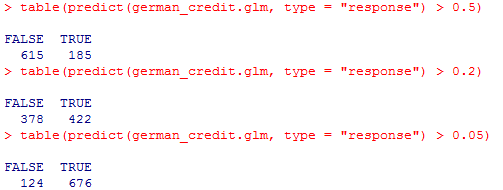
German\_credit.glm has 14 variables



Prediction (using probaility)

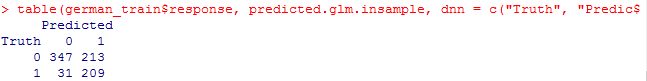


Cutoffs



In-sample

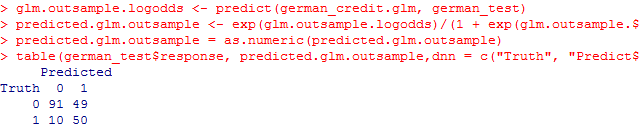




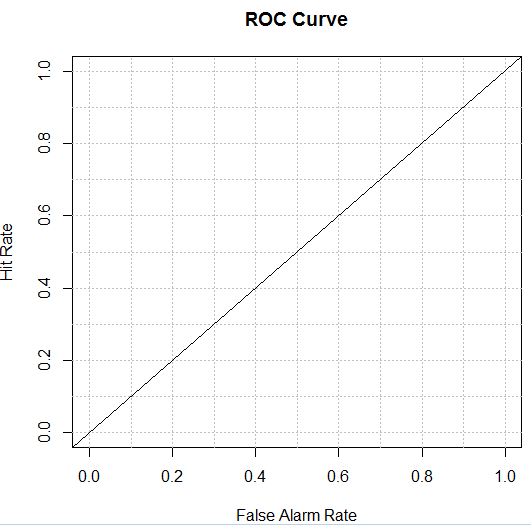
Error rate



Out-sample



ROC curve

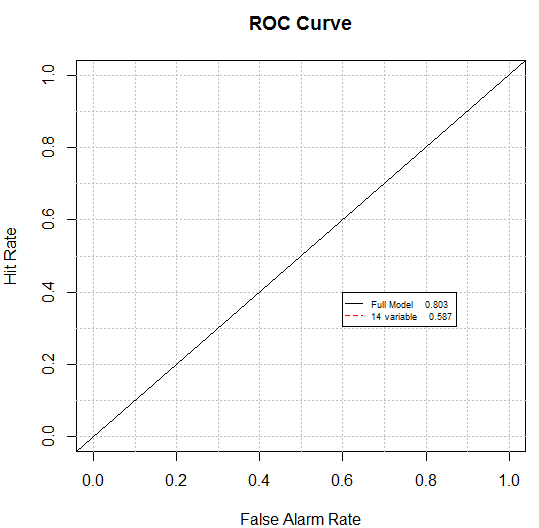


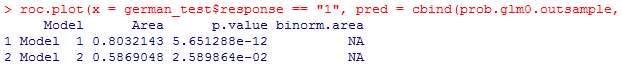


Out sample error rate

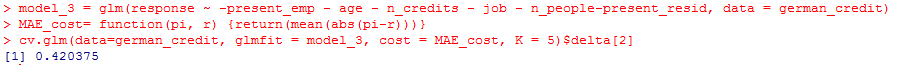


Comparing both models (ROC curve of both with area)





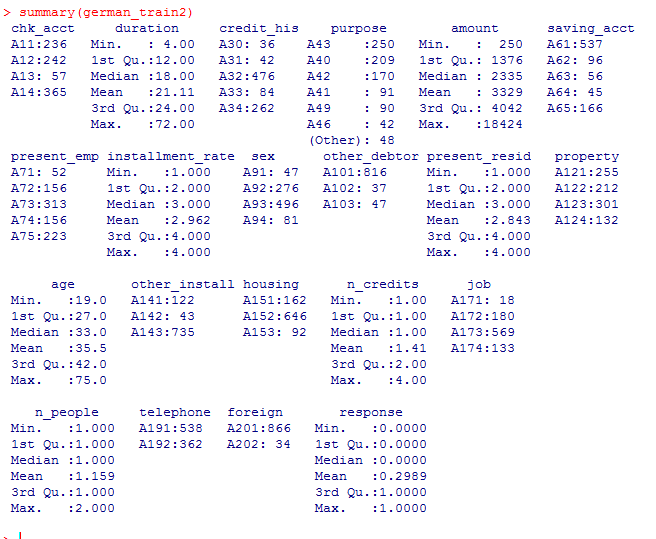
5 fold cross validation



<http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)>

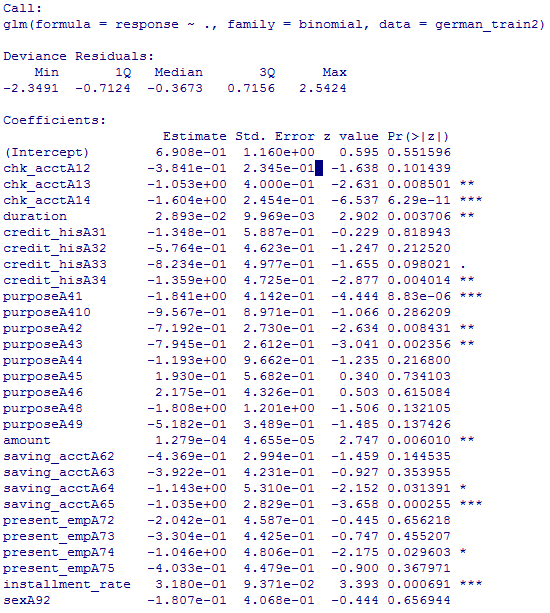
90 Percent

Summary



Logistic model on 90%

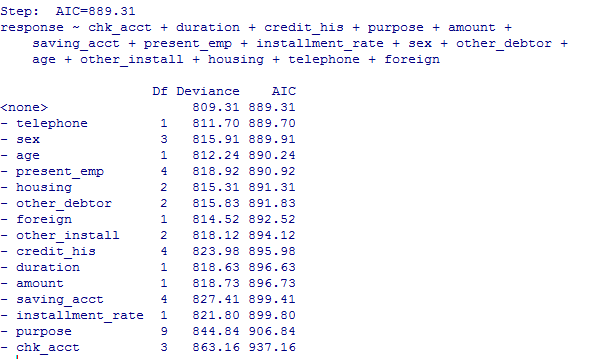






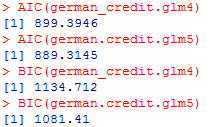
Step wise





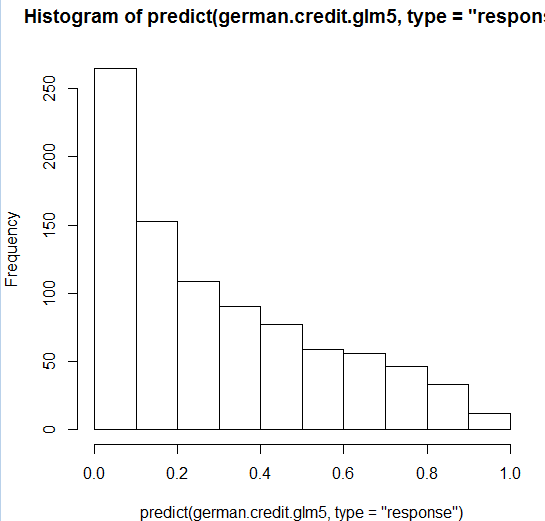
German\_credit.glm0 has all variables

German\_credit.glm has 15 variables

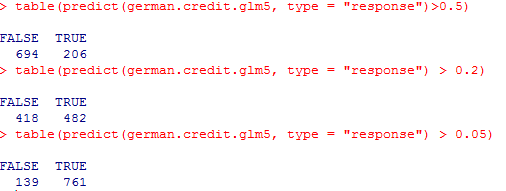


Probability



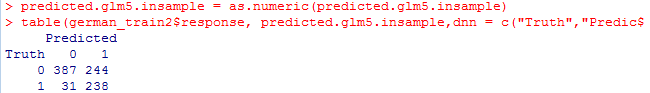


Cutoff



In sample

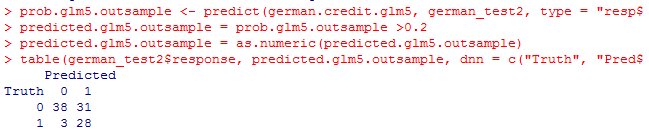




Error rate



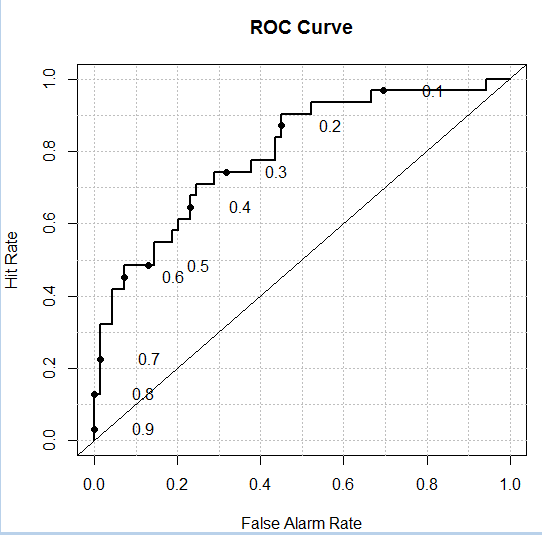
Out of sample



Error



ROC curve





5 fold cross validation

