Q2: Financial Condition of Banks

G Park

2022-12-04

# **Question Two**

The file banks.csv includes data on a sample of 20 banks. The “Financial Condition” column records the judgment of an expert on the financial condition of each bank. This outcome variable takes on of two possible values –weak (1) or strong (0)– according to the financial condition of the bank. The predictors are two ratios used in the financial analysis of banks: TotLns&Lses/Assets is the ratio of total loans and leases to total assets and TotExp/Assets is the ratio of total expenses to total assets. The target is to use the two ratios for classifying the financial condition of a new bank.

## 2.1 Partition records into 60% for training and 40% for validation sets. Then fit a logistic regression to Financial\_Condition as function of TotLns&Lses/Assets and TotExp/Assets on training sets and show summarized logistic results. (10 points)

I have used the tidymodels initial\_split, training and testing functions to split the data into training and validation sets. I then use the glm function to fit a logistoc regression and summarise the results.

##   
## Call:  
## glm(formula = Financial\_Condition ~ . - Obs, family = binomial(link = "logit"),   
## data = bank\_training\_set)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0309 -0.4287 -0.0369 0.6630 1.2100   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -10.22 5.89 -1.74 0.083 .  
## `TotExp/Assets` 57.07 52.85 1.08 0.280   
## `TotLns&Lses/Assets` 6.62 5.44 1.22 0.224   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 16.6355 on 11 degrees of freedom  
## Residual deviance: 9.9891 on 9 degrees of freedom  
## AIC: 15.99  
##   
## Number of Fisher Scoring iterations: 5

## 2.2 Now based on the logistic regression results, make predicted probability for validation set and show their predicted probability. (10 points)

I make predictions on the validation set and print the predicted probabilities below.

## 1 2 3 4 5 6 7 8   
## 0.718 0.652 0.797 0.936 0.160 0.425 0.061 0.093

## Now, create a data frame for the first 5 actual records of validation sets along with their predicted probability. (10 points)

Using a cutoff of 50% (0.5), I make show the predictions from the model. If probability is at least 0.5, I predict 1, else I predict 0.

## # A tibble: 5 × 3  
## Financial\_Condition predicted\_probs predicted\_outcomes  
## <dbl> <dbl> <dbl>  
## 1 1 0.718 1  
## 2 1 0.652 1  
## 3 1 0.797 1  
## 4 1 0.936 1  
## 5 0 0.160 0