

Financial Condition of Banks

G Park

2022-12-02

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 logistic 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
## -2.0834  -0.4359  -0.0482   0.7231   1.1925
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -10.61      5.87   -1.81   0.071 .
## `TotExp/Assets`    55.56     52.24    1.06   0.287
## `TotLns&Lses/Assets`  7.49      8.63    0.87   0.386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 16.636  on 11  degrees of freedom
## Residual deviance: 10.150  on  9  degrees of freedom
## AIC: 16.15
##
## 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
## 0.716 0.884 0.661 0.928 0.175 0.465 0.034 0.054
```

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 x 3
##   Financial_Condition predicted_probs predicted_outcomes
##           <dbl>           <dbl>           <dbl>
## 1                1         0.716                1
## 2                1         0.884                1
## 3                1         0.661                1
## 4                1         0.928                1
## 5                0         0.175                0
```