Assignment7

Yiming Ge

11/7/2020

Problem 1 Student Application Data read the data

```
url <- 'https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/Univ%20Admissions.c
studentAppl <- read.csv(url,header=TRUE, stringsAsFactors=TRUE)
dim(studentAppl)
```

```
## [1] 225015     10
```

```
names(studentAppl)
```

```
##  [1] "x.Country"                "x.State"
##  [3] "x.Gender"                 "x.Source"
##  [5] "x.GPA"                    "x.SAT_Score"
##  [7] "x.DistancetoCampus_miles" "x.HouseholdIncome"
##  [9] "x.Status.1"               "x.InState"
```

```
head(studentAppl)
```

```
##   x.Country x.State x.Gender             x.Source x.GPA x.SAT_Score
## 1       USA      NY     Male NRCCUA-PurchaseNames     2
## 2       USA      NY     Male NRCCUA-PurchaseNames     2
## 3       USA      NY   Female NRCCUA-PurchaseNames     2
## 4       USA      FL     Male NRCCUA-PurchaseNames     2
## 5       USA      NJ   Female NRCCUA-PurchaseNames     2
## 6       USA      NJ   Female NRCCUA-PurchaseNames     2
##   x.DistancetoCampus_miles x.HouseholdIncome x.Status.1 x.InState
## 1                 44.54265             36990    SUSPECT         N
## 2                 40.50179             33919    SUSPECT         N
## 3                211.00019             55624    SUSPECT         N
## 4               1013.99259             33105    SUSPECT         N
## 5                 59.80009             25999    SUSPECT         N
## 6                 64.48530             41162    SUSPECT         N
```

clean the data

```
studentAppl$x.SAT_Score[studentAppl$x.SAT_Score == '']<-NA
#data cleaning
SA_noNA <- studentAppl[!is.infinite(studentAppl$x.HouseholdIncome),] #remove the infinite row in househ
#replace na with mean
```

```
SA_noNA$x.GPA[is.na(SA_noNA$x.GPA)] <- mean(na.omit(SA_noNA$x.GPA))

SA_noNA$x.DistancetoCampus_miles[is.na(SA_noNA$x.DistancetoCampus_miles)]<-
  mean(na.omit(SA_noNA$x.DistancetoCampus_miles))
#based on status.1, applicant and prospect are 1, suspect is 0
SA_noNA$x.Status.1<-ifelse(SA_noNA$x.Status.1 %in% ’SUSPECT’,0,1)
#based on InState, yes is 1, no is 0
SA_noNA$x.InState<-ifelse(SA_noNA$x.InState %in% ’N’,0,1)
summary(SA_noNA)
```

```
##  x.Country       x.State        x.Gender
##     :    77   NY     :66426         : 20150
##  USA:224930   NJ     :54054   Female:113142
##               CT     :39602   Male  : 91715
##               MA     :20081
##               MD     :18177
##               IL     :14345
##               (Other):12322
##                          x.Source         x.GPA            x.SAT_Score
##  NRCCUA-PurchaseNames         :67504   Min.   :2.000   930 - 1070 : 17754
##  CollegeBoard-Senior_Search   :54224   1st Qu.:3.000   1080 - 1350: 14872
##  CollegeBoard-Juniors_Search  :50921   Median :3.105   1360 - 1530:  2024
##  CollegeBoard-Sophomore_Search:17372   Mean   :3.105   930 - 980  :   156
##  ACT-Other                    : 7953   3rd Qu.:3.105   990 - 1040 :     8
##  CollegeBoard-Other           : 7493   Max.   :4.000   (Other)    :    25
##  (Other)                      :19540                   NA’s       :190168
##  x.DistancetoCampus_miles x.HouseholdIncome      x.Status.1       x.InState
##  Min.   :   0.052         Min.   : 2.004e+04   Min.   :0.0000   Min.   :0.000
##  1st Qu.:  43.556         1st Qu.: 5.689e+04   1st Qu.:0.0000   1st Qu.:0.000
##  Median :  62.509         Median : 8.725e+04   Median :0.0000   Median :0.000
##  Mean   : 144.431         Mean   :2.600e+143   Mean   :0.1095   Mean   :0.176
##  3rd Qu.: 139.472         3rd Qu.: 1.149e+05   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :5001.483         Max.   :5.850e+148   Max.   :1.0000   Max.   :1.000
##
```

train test split

```
set.seed(88) # setting seed to reproduce results of random sampling
split<-(.75)
trainingRowIndex <- sample(1:nrow(SA_noNA),(split)*nrow(SA_noNA)) # row indices for training data
SAtrainingData <- SA_noNA[trainingRowIndex, ] # model training data
SAtestData <- SA_noNA[-trainingRowIndex, ] # test data
```

develop the model

```
# Model
model<-{x.Status.1 ~ .}
SA.lm <- glm(model, data=SAtrainingData, family = binomial(link = "logit")) # build the model
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# Review diagnostic measures
summary(SA.lm)
```

```
##
## Call:
## glm(formula = model, family = binomial(link = "logit"), data = SAtrainingData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9429  -0.2957  -0.2424  -0.1962   3.2133
##
## Coefficients: (2 not defined because of singularities)
##                                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        -9.249e+00  3.788e+03  -0.002   0.9981
## x.CountryUSA                        1.297e+01  1.682e+03   0.008   0.9938
## x.StateAZ                          -1.576e+01  2.400e+03  -0.007   0.9948
## x.StateCA                          -1.643e+01  2.626e+02  -0.063   0.9501
## x.StateCO                          -1.548e+01  2.400e+03  -0.006   0.9949
## x.StateCT                           1.192e+00  6.509e-01   1.831   0.0671 .
## x.StateDC                          -1.262e+01  9.646e+02  -0.013   0.9896
## x.StateFL                          -1.398e+01  7.901e+02  -0.018   0.9859
## x.StateGA                          -1.405e+01  1.071e+03  -0.013   0.9895
## x.StateIL                          -9.360e-01  1.059e+00  -0.884   0.3767
## x.StateIN                          -1.350e+01  2.400e+03  -0.006   0.9955
## x.StateLA                          -1.416e+01  1.697e+03  -0.008   0.9933
## x.StateMA                           1.640e-01  6.143e-01   0.267   0.7895
## x.StateMD                           8.055e-03  5.953e-01   0.014   0.9892
## x.StateME                          -1.295e+01  1.687e+03  -0.008   0.9939
## x.StateMO                          -1.445e+01  2.400e+03  -0.006   0.9952
## x.StateMT                          -1.544e+01  2.400e+03  -0.006   0.9949
## x.StateNC                          -1.326e+01  1.383e+03  -0.010   0.9924
## x.StateNE                          -1.401e+01  2.400e+03  -0.006   0.9953
## x.StateNH                           2.359e-01  6.287e-01   0.375   0.7075
## x.StateNJ                           5.096e-01  6.255e-01   0.815   0.4153
## x.StateNV                          -1.583e+01  2.400e+03  -0.007   0.9947
## x.StateNY                           7.192e-01  6.407e-01   1.123   0.2616
## x.StatePA                          -2.712e-01  6.548e-01  -0.414   0.6788
## x.StateRI                          -2.091e-01  7.628e-01  -0.274   0.7840
## x.StateTX                          -1.491e+01  9.710e+02  -0.015   0.9878
## x.StateVA                          -6.838e-01  9.247e-01  -0.739   0.4596
## x.StateVT                                  NA         NA      NA       NA
## x.GenderFemale                     -2.189e+00  1.181e+00  -1.854   0.0637 .
## x.GenderMale                       -2.410e+00  1.181e+00  -2.041   0.0413 *
## x.SourceCollegeBoard-Other          1.178e+01  2.400e+03   0.005   0.9961
## x.SourceCollegeBoard-Senior_Search  1.241e+01  2.400e+03   0.005   0.9959
## x.SourceNRCCUA-Other               -1.481e+00  2.939e+03  -0.001   0.9996
## x.SourceNRCCUA-Senior_Search       -1.312e+00  2.497e+03  -0.001   0.9996
## x.SourceProspects-Senior_Search     4.242e+01  2.762e+03   0.015   0.9877
## x.GPA                               1.100e+00  1.481e+00   0.743   0.4577
## x.SAT_Score1080 - 1350             -2.079e+01  2.400e+03  -0.009   0.9931
## x.SAT_Score1110 - 1160             -3.430e+01  2.646e+03  -0.013   0.9897
## x.SAT_Score1170 - 1220             -3.464e+01  2.722e+03  -0.013   0.9898
## x.SAT_Score1230 - 1280             -3.378e+01  2.561e+03  -0.013   0.9895
## x.SAT_Score1290 - 1340             -3.341e+01  2.838e+03  -0.012   0.9906
## x.SAT_Score1350 - 1400             -3.380e+01  2.816e+03  -0.012   0.9904
## x.SAT_Score1360 - 1530             -2.070e+01  2.400e+03  -0.009   0.9931
## x.SAT_Score1410 - 1460             -3.513e+01  3.393e+03  -0.010   0.9917
## x.SAT_Score930 - 1070              -2.031e+01  2.400e+03  -0.008   0.9932
## x.SAT_Score930 - 980               -2.174e+01  2.400e+03  -0.009   0.9928
## x.SAT_Score990 - 1040              -1.833e+01  2.400e+03  -0.008   0.9939
## x.DistancetoCampus_miles            1.580e-03  1.474e-03   1.072   0.2837
## x.HouseholdIncome                  -6.268e-06  1.016e-06  -6.170 6.85e-10 ***
## x.InState                                  NA         NA      NA       NA
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7814.9  on 26173  degrees of freedom
## Residual deviance: 7521.6  on 26126  degrees of freedom
##   (142581 observations deleted due to missingness)
## AIC: 7617.6
##
## Number of Fisher Scoring iterations: 15
```

pick features to build new model

```
model1<-{x.Status.1 ~ x.Gender + x.Source + x.GPA + x.DistancetoCampus_miles + x.HouseholdIncome}
SA.lm1 <- glm(model1, data=SAtrainingData, family = binomial(link = "logit"))
summary(SA.lm1)
```

```
##
## Call:
## glm(formula = model1, family = binomial(link = "logit"), data = SAtrainingData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8614  -0.2850  -0.2464  -0.2050   3.7805
##
## Coefficients:
##                                          Estimate Std. Error  z value Pr(>|z|)
## (Intercept)                            2.861e-01  2.222e-01    1.288 0.197881
## x.GenderFemale                        -3.281e+00  3.208e-02 -102.275  < 2e-16
## x.GenderMale                          -3.620e+00  3.538e-02 -102.310  < 2e-16
## x.SourceACT-No Data                    4.071e+00  7.506e-01    5.424 5.84e-08
## x.SourceACT-Other                     -1.533e+00  2.238e-01   -6.849 7.42e-12
## x.SourceACT-Senior_Search              1.648e+00  2.156e-01    7.643 2.12e-14
## x.SourceCollegeBoard-Juniors_Search   -2.340e-01  1.840e-01   -1.272 0.203298
## x.SourceCollegeBoard-No Data           3.649e+00  2.689e-01   13.572  < 2e-16
## x.SourceCollegeBoard-Other            -2.748e-01  1.944e-01   -1.413 0.157573
## x.SourceCollegeBoard-Prospects         1.880e+01  8.625e+02    0.022 0.982608
## x.SourceCollegeBoard-PurchaseNames     5.205e-01  4.866e-01    1.070 0.284827
## x.SourceCollegeBoard-Senior_Search    -6.290e-01  1.850e-01   -3.400 0.000673
## x.SourceCollegeBoard-Sophomore_Search  4.919e-01  1.848e-01    2.662 0.007771
## x.SourceNo Data                        1.949e+01  5.611e+01    0.347 0.728368
## x.SourceNRCCUA-Juniors_Search          1.689e+00  1.046e+00    1.614 0.106427
## x.SourceNRCCUA-No Data                -9.490e-01  1.887e-01   -5.030 4.90e-07
## x.SourceNRCCUA-Other                  -5.251e-01  1.173e+00   -0.448 0.654347
## x.SourceNRCCUA-Prospects               5.005e+00  1.075e+00    4.658 3.20e-06
## x.SourceNRCCUA-PurchaseNames          -4.191e-01  1.834e-01   -2.285 0.022323
## x.SourceNRCCUA-Senior_Search          -1.361e+01  7.982e+02   -0.017 0.986393
## x.SourceNRCCUA-Sophomore_Search        3.704e+00  1.242e+00    2.984 0.002847
## x.SourceProspects-Juniors_Search       2.103e+01  9.226e+02    0.023 0.981812
## x.SourceProspects-No Data              6.190e+00  4.710e-01   13.142  < 2e-16
## x.SourceProspects-Other                1.952e+01  7.572e+02    0.026 0.979433
## x.SourceProspects-Prospects            6.882e+00  2.250e-01   30.582  < 2e-16
## x.SourceProspects-PurchaseNames        5.706e+00  1.047e+00    5.450 5.05e-08
## x.SourceProspects-Senior_Search        1.934e+01  9.794e+02    0.020 0.984245
## x.SourceProspects-Sophomore_Search     1.942e+01  2.400e+03    0.008 0.993542
## x.SourcePurchaseNames-No Data          1.605e+01  1.200e+03    0.013 0.989327
## x.SourcePurchaseNames-Other            1.926e+01  2.400e+03    0.008 0.993595
## x.SourcePurchaseNames-PurchaseNames   -3.147e+00  2.311e-01  -13.617  < 2e-16
## x.GPA                                  9.289e-02  4.098e-02    2.267 0.023412
## x.DistancetoCampus_miles              -1.466e-03  6.926e-05  -21.166  < 2e-16
## x.HouseholdIncome                      2.067e-49  2.680e-47    0.008 0.993846
##
## (Intercept)
## x.GenderFemale                        ***
## x.GenderMale                          ***
## x.SourceACT-No Data                   ***
## x.SourceACT-Other                     ***
## x.SourceACT-Senior_Search             ***
## x.SourceCollegeBoard-Juniors_Search
## x.SourceCollegeBoard-No Data          ***
## x.SourceCollegeBoard-Other
## x.SourceCollegeBoard-Prospects
## x.SourceCollegeBoard-PurchaseNames
## x.SourceCollegeBoard-Senior_Search    ***
## x.SourceCollegeBoard-Sophomore_Search **
## x.SourceNo Data
## x.SourceNRCCUA-Juniors_Search
## x.SourceNRCCUA-No Data                ***
## x.SourceNRCCUA-Other
## x.SourceNRCCUA-Prospects              ***
## x.SourceNRCCUA-PurchaseNames          *
## x.SourceNRCCUA-Senior_Search
## x.SourceNRCCUA-Sophomore_Search       **
## x.SourceProspects-Juniors_Search
## x.SourceProspects-No Data             ***
## x.SourceProspects-Other
## x.SourceProspects-Prospects           ***
## x.SourceProspects-PurchaseNames       ***
## x.SourceProspects-Senior_Search
## x.SourceProspects-Sophomore_Search
## x.SourcePurchaseNames-No Data
## x.SourcePurchaseNames-Other
## x.SourcePurchaseNames-PurchaseNames   ***
## x.GPA                                 *
## x.DistancetoCampus_miles              ***
## x.HouseholdIncome
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 116450  on 168754  degrees of freedom
## Residual deviance:  57287  on 168721  degrees of freedom
## AIC: 57355
##
## Number of Fisher Scoring iterations: 15
```

test prediction accuracy

```
SAresponse<- ifelse(predict(SA.lm1, SAtestData, type = "response")>.5, 1, 0) # predict distance
SAactuals_preds <- data.frame(cbind(actuals=SAtestData$x.Status.1, predicted=SAresponse))
head(SAactuals_preds)
```

```
##    actuals predicted
## 2        0         0
## 3        0         0
## 6        0         0
## 8        0         0
## 9        0         0
## 11       0         0
```

```
# simple correlation between actuals vs predicted is an accuracy measure.
# a higher correlation accuracy implies similar directional movement
SAcorrelation_accuracy <- cor(SAactuals_preds)
SAcorrelation_accuracy
```

```
##             actuals predicted
## actuals   1.0000000 0.7071838
## predicted 0.7071838 1.0000000
```

Conclusion and recommendation: Based on the high correlation between the predicted and actual values, I can say that my binary model between likeness to apply and gender, source, GPA, distance to campus miles, household income has good performance. Based on the coefficient of these variables, I can give a recommendation that the school should message more to the female student with high gpa, house income and less distance to campus miles with source from prospects.

Problem2: Bank Marketing read the data

```
url <- ’https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/bank_marketing.csv’
bankMarket <- read.csv(url,header=TRUE, stringsAsFactors=TRUE)
bankMarket$y<-ifelse(bankMarket$y %in% ’no’,0,1)
dim(bankMarket)
```

```
## [1] 41188    21
```

```
names(bankMarket)
```

```
##  [1] "age"            "job"            "marital"        "education"
##  [5] "default"        "housing"        "loan"           "contact"
##  [9] "month"          "day_of_week"    "duration"       "campaign"
## [13] "pdays"          "previous"       "poutcome"       "emp.var.rate"
## [17] "cons.price.idx" "cons.conf.idx"  "euribor3m"      "nr.employed"
## [21] "y"
```

```
head(bankMarket)
```

```
##   age       job marital   education default housing loan   contact month
## 1  56 housemaid married    basic.4y      no      no   no telephone   may
## 2  57  services married high.school unknown      no   no telephone   may
## 3  37  services married high.school      no     yes   no telephone   may
## 4  40    admin. married    basic.6y      no      no   no telephone   may
## 5  56  services married high.school      no      no  yes telephone   may
## 6  45  services married    basic.9y unknown      no   no telephone   may
##   day_of_week duration campaign pdays previous    poutcome emp.var.rate
## 1         mon      261        1   999        0 nonexistent          1.1
## 2         mon      149        1   999        0 nonexistent          1.1
## 3         mon      226        1   999        0 nonexistent          1.1
## 4         mon      151        1   999        0 nonexistent          1.1
## 5         mon      307        1   999        0 nonexistent          1.1
## 6         mon      198        1   999        0 nonexistent          1.1
##   cons.price.idx cons.conf.idx euribor3m nr.employed y
## 1         93.994         -36.4     4.857        5191 0
## 2         93.994         -36.4     4.857        5191 0
## 3         93.994         -36.4     4.857        5191 0
## 4         93.994         -36.4     4.857        5191 0
## 5         93.994         -36.4     4.857        5191 0
## 6         93.994         -36.4     4.857        5191 0
```

train test split

```
set.seed(43) # setting seed to reproduce results of random sampling
split<-(.8)
trainingRowIndex <- sample(1:nrow(bankMarket),(split)*nrow(bankMarket)) # row indices for training data
BMtrainingData <- bankMarket[trainingRowIndex, ] # model training data
BMtestData <- bankMarket[-trainingRowIndex, ] # test data
```

develop the model

```
# Model
model<-{y ~ .}
BM.lm <- glm(model, data=BMtrainingData, family = binomial(link = "logit")) # build the model
# Review diagnostic measures
summary(BM.lm)
```

```
##
## Call:
## glm(formula = model, family = binomial(link = "logit"), data = BMtrainingData)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -5.9876  -0.2955  -0.1858  -0.1333   3.4140
##
## Coefficients: (1 not defined because of singularities)
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -2.752e+02  4.340e+01  -6.342 2.27e-10 ***
## age                           1.330e-03  2.746e-03   0.484  0.62808
## jobblue-collar               -2.656e-01  8.958e-02  -2.965  0.00303 **
## jobentrepreneur              -1.265e-01  1.405e-01  -0.901  0.36779
## jobhousemaid                  1.181e-01  1.625e-01   0.727  0.46730
## jobmanagement                -2.013e-02  9.543e-02  -0.211  0.83295
## jobretired                    3.256e-01  1.209e-01   2.694  0.00707 **
## jobself-employed             -8.893e-02  1.301e-01  -0.684  0.49412
## jobservices                  -1.463e-01  9.793e-02  -1.494  0.13508
## jobstudent                    2.702e-01  1.246e-01   2.169  0.03006 *
## jobtechnician                -3.934e-02  7.991e-02  -0.492  0.62254
## jobunemployed                 1.125e-01  1.412e-01   0.797  0.42548
## jobunknown                   -9.114e-02  2.680e-01  -0.340  0.73378
## maritalmarried                1.017e-02  7.681e-02   0.132  0.89471
## maritalsingle                 9.862e-02  8.781e-02   1.123  0.26142
## maritalunknown               -5.925e-02  4.684e-01  -0.127  0.89933
## educationbasic.6y             1.074e-01  1.355e-01   0.792  0.42816
## educationbasic.9y             4.567e-02  1.066e-01   0.428  0.66830
## educationhigh.school          3.318e-02  1.033e-01   0.321  0.74809
## educationilliterate           1.062e+00  8.111e-01   1.310  0.19027
## educationprofessional.course  1.075e-01  1.135e-01   0.947  0.34387
## educationuniversity.degree    2.360e-01  1.032e-01   2.286  0.02223 *
## educationunknown              1.929e-01  1.332e-01   1.448  0.14762
## defaultunknown               -2.656e-01  7.488e-02  -3.547  0.00039 ***
## defaultyes                   -7.383e+00  1.391e+02  -0.053  0.95767
## housingunknown               -1.873e-01  1.610e-01  -1.163  0.24482
## housingyes                    1.325e-02  4.645e-02   0.285  0.77545
## loanunknown                          NA         NA      NA       NA
## loanyes                      -5.846e-02  6.449e-02  -0.906  0.36472
## contacttelephone             -7.165e-01  8.806e-02  -8.136 4.08e-16 ***
## monthaug                      9.535e-01  1.360e-01   7.010 2.39e-12 ***
## monthdec                      5.113e-01  2.368e-01   2.159  0.03087 *
## monthjul                      1.680e-01  1.090e-01   1.541  0.12320
## monthjun                     -5.921e-01  1.441e-01  -4.109 3.97e-05 ***
## monthmar                      2.171e+00  1.625e-01  13.358  < 2e-16 ***
## monthmay                     -4.000e-01  9.345e-02  -4.281 1.86e-05 ***
## monthnov                     -3.189e-01  1.354e-01  -2.355  0.01851 *
## monthoct                      2.827e-01  1.738e-01   1.626  0.10389
## monthsep                      4.072e-01  2.054e-01   1.983  0.04740 *
## day_of_weekmon               -1.050e-01  7.462e-02  -1.407  0.15947
## day_of_weekthu                8.080e-02  7.213e-02   1.120  0.26267
## day_of_weektue                1.374e-01  7.419e-02   1.852  0.06396 .
## day_of_weekwed                1.937e-01  7.415e-02   2.612  0.00899 **
## duration                      4.702e-03  8.268e-05  56.873  < 2e-16 ***
## campaign                     -3.196e-02  1.255e-02  -2.546  0.01090 *
## pdays                        -7.470e-04  2.446e-04  -3.054  0.00226 **
## previous                      8.669e-04  6.692e-02   0.013  0.98967
## poutcomenonexistent           4.306e-01  1.067e-01   4.035 5.47e-05 ***
## poutcomesuccess               1.150e+00  2.388e-01   4.817 1.46e-06 ***
## emp.var.rate                 -1.895e+00  1.617e-01 -11.717  < 2e-16 ***
## cons.price.idx                2.432e+00  2.861e-01   8.502  < 2e-16 ***
## cons.conf.idx                 2.268e-02  8.741e-03   2.595  0.00946 **
## euribor3m                     2.986e-01  1.461e-01   2.044  0.04091 *
## nr.employed                   8.475e-03  3.522e-03   2.407  0.01610 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 23091  on 32949  degrees of freedom
## Residual deviance: 13556  on 32897  degrees of freedom
## AIC: 13662
##
## Number of Fisher Scoring iterations: 10
```

test prediction accuracy

```
BMresponse<- ifelse(predict(BM.lm, BMtestData, type = "response")>.5, 1, 0) # predict distance
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

```
BMactuals_preds <- data.frame(cbind(actuals=BMtestData$y, predicted=BMresponse))
head(BMactuals_preds)
```

```
##    actuals predicted
## 11       0         0
## 22       0         0
## 27       0         0
## 28       0         0
## 31       0         0
## 37       0         0
```

```
# simple correlation between actuals vs predicted is an accuracy measure.
# a higher correlation accuracy implies similar directional movement
BMcorrelation_accuracy <- cor(BMactuals_preds)
BMcorrelation_accuracy
```

```
##             actuals predicted
## actuals   1.0000000 0.4578328
## predicted 0.4578328 1.0000000
```

Conclusion and recommendation: Based on the correlation between the predicted and actual values, I can say that although my binary model between whether bank term deposit would be ('yes') or not ('no') subscribed between "age""job""marital""education""default""housing""loan""contact""month""day_of_week""duration""campaign""p "nr.employed" does not have very good performance, based on the coefficient of these variables, I still can give a recommendation that the BANK should implement marketing campaigns and focus more on retired people and student with university degree on march, august, september and december.