exam-2-13

Pradeep Paladugula

7/17/2020

```r
library('tidyverse')
```

```
## — Attaching packages ─────────────────────────────────────────
──────────────────────── tidyverse 1.3.0 —
## ✓ ggplot2 3.3.0     ✓ purrr   0.3.3
## ✓ tibble  2.1.3     ✓ dplyr   0.8.5
## ✓ tidyr   1.1.0     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.4.0
## — Conflicts ──────────────────────────────────────────────────
──────────────────────── tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library("readxl")
appData2 <- read_excel('Application2.xlsx')
appData2
```

```
## # A tibble: 2,160 x 3
##    Major      Degree     Condition
##    <chr>      <chr>      <chr>
##  1 Math       HighSchool High
##  2 Education  Bachelor   Low
##  3 Business   Master     High
##  4 Education  HighSchool High
##  5 Business   Bachelor   Low
##  6 Math       Master     High
##  7 Education  HighSchool High
##  8 Math       Bachelor   Low
##  9 Business   Master     Low
## 10 Business   HighSchool High
## # … with 2,150 more rows
```

```r
with(appData2, table(Major, Degree, Condition))
```

```
## , , Condition = High
##
##           Degree
## Major      Bachelor HighSchool Master
##   Business      183        121    192
```

```
##    Education      158         194     100
##    Math           131         115     238
##
## , , Condition = Low
##
##            Degree
## Major        Bachelor HighSchool Master
##    Business         85          80     59
##    Education        93         141     34
##    Math             70          68     97

appData2$Major <- factor(appData2$Major)
appData2$Degree <- factor(appData2$Degree)
appData2$Condition <- factor(appData2$Condition)

model1 <- glm(Condition~Major, data = appData2, family = "binomial")
model1

##
## Call:  glm(formula = Condition ~ Major, family = "binomial", data =
appData2)
##
## Coefficients:
##    (Intercept)  MajorEducation       MajorMath
##       -0.79493         0.27223         0.07243
##
## Degrees of Freedom: 2158 Total (i.e. Null);   2156 Residual
##    (1 observation deleted due to missingness)
## Null Deviance:        2759
## Residual Deviance: 2752  AIC: 2758

summary(model1)

##
## Call:
## glm(formula = Condition ~ Major, family = "binomial", data = appData2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9649  -0.8897  -0.8633   1.4059   1.5281
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.79493    0.08050  -9.875   <2e-16 ***
## MajorEducation  0.27223    0.11146   2.442   0.0146 *
## MajorMath       0.07243    0.11315   0.640   0.5221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 2758.5  on 2158  degrees of freedom
## Residual deviance: 2752.1  on 2156  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 2758.1
##
## Number of Fisher Scoring iterations: 4

model2 <- glm(Condition~Major+Degree, data = appData2, family = "binomial")
model2

##
## Call:  glm(formula = Condition ~ Major + Degree, family = "binomial",
##     data = appData2)
##
## Coefficients:
##     (Intercept)    MajorEducation        MajorMath  DegreeHighSchool
##         -0.7388            0.1727           0.1225            0.2293
##     DegreeMaster
##         -0.3772
##
## Degrees of Freedom: 2158 Total (i.e. Null);   2154 Residual
##   (1 observation deleted due to missingness)
## Null Deviance:       2759
## Residual Deviance: 2725  AIC: 2735

summary(model2)

##
## Call:
## glm(formula = Condition ~ Major + Degree, family = "binomial",
##     data = appData2)
##
## Deviance Residuals:
##     Min       1Q    Median        3Q       Max
## -1.0381  -0.9483   -0.7937    1.3456    1.6730
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.7388     0.1007  -7.336  2.2e-13 ***
## MajorEducation     0.1727     0.1137   1.519  0.12883
## MajorMath          0.1225     0.1146   1.069  0.28524
## DegreeHighSchool   0.2293     0.1100   2.085  0.03704 *
## DegreeMaster      -0.3772     0.1172  -3.218  0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2758.5  on 2158  degrees of freedom
## Residual deviance: 2724.7  on 2154  degrees of freedom
##   (1 observation deleted due to missingness)
```

```
## AIC: 2734.7
##
## Number of Fisher Scoring iterations: 4

#as modal cantains two or more terms
#vif(model2)

modelchi1 <- model1$null.deviance-model1$deviance
modelchi1

## [1] 6.456158

cdf1 <- model1$df.null-model1$df.residual
cdf1

## [1] 2

chisqp1 <- 1 - pchisq(modelchi1, cdf1)
chisqp1

## [1] 0.03963355

modelchi2 <- model2$null.deviance-model2$deviance
modelchi2

## [1] 33.78481

cdf2 <- model2$df.null-model2$df.residual
cdf2

## [1] 4

chisqp2 <- 1 - pchisq(modelchi2, cdf2)
chisqp2

## [1] 8.248783e-07

anova(model1, model2)

## Analysis of Deviance Table
##
## Model 1: Condition ~ Major
## Model 2: Condition ~ Major + Degree
##   Resid. Df Resid. Dev Df Deviance
## 1      2156     2752.1
## 2      2154     2724.7  2   27.329

anova(model1, model2, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Condition ~ Major
## Model 2: Condition ~ Major + Degree
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
```

```
## 1      2156     2752.1
## 2      2154     2724.7  2   27.329 1.163e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R2.hl1 <- modelchi1/model1$null.deviance
R2.hl1

## [1] 0.002340441

R2.hl2 <- modelchi2/model2$null.deviance
R2.hl2

## [1] 0.01224743

model1$coefficients

##    (Intercept) MajorEducation      MajorMath
##    -0.79492987     0.27223468     0.07243048

exp(model1$coefficients)

##    (Intercept) MajorEducation      MajorMath
##      0.4516129      1.3128951      1.0751181

exp(confint(model1))

## Waiting for profiling to be done...

##                    2.5 %    97.5 %
## (Intercept)    0.3850562 0.5280124
## MajorEducation 1.0555389 1.6341525
## MajorMath      0.8613037 1.3422836

model2$coefficients

##      (Intercept)   MajorEducation      MajorMath DegreeHighSchool
##       -0.7388468        0.1727225      0.1224802        0.2293032
##     DegreeMaster
##       -0.3772254

exp(model2$coefficients)

##      (Intercept)   MajorEducation      MajorMath DegreeHighSchool
##        0.4776644        1.1885362      1.1302968        1.2577233
##     DegreeMaster
##        0.6857615

exp(confint(model2))

## Waiting for profiling to be done...

##                    2.5 %    97.5 %
## (Intercept)      0.3914197 0.5810239
```

```
## MajorEducation    0.9511657 1.4857129
## MajorMath         0.9029916 1.4153957
## DegreeHighSchool 1.0140625 1.5607040
## DegreeMaster      0.5446313 0.8624354
```

Summary: A logistic regression model was conducted to predict whether an individual's starting Condition(salary) is low or high depend on major and degree. Both intercept(b= -0.7388, z value = -7.336, p-value = <0.001), MajorEducation (b = 0.1727,z-avlue = 1.519, p = 0.12883(>0.05)) and MajorMath (b= 0.1225, z-value, 1.069, p-value = 0.28524 (>0.05)) which are not statistically significant. The odds ratio of Education is 0.0.95, suggesting people studies education have a 0.95 times of chance of receive high starting salary than those who didn't. The odds ratio of Math is 0.90, suggesting people studies Math have a 0.90 times of chance of receive high starting salary than those who didn't. The odds ratio of highschooler is 1.014, suggesting people has a high school degree have a 00.14 times of change of receive high starting salary. The odds ratio of Master's degree is 0.54, suggesting people has a Master's degree have a 0.54 times of change of receive high starting salary. Because DegreeHighSchool(b =0.2293, x-value =2.085, p-value=0.03704) and DegreeMaster(b =-0.3772, z-value = -3.218, p-value = 0.00129) were statistical significant.