UAS-ADK-2021.R

Ridson Alfarizal

2021-12-24

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.6 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

## Warning: package 'tibble' was built under R version 4.0.3

## Warning: package 'tidyr' was built under R version 4.0.3

## Warning: package 'readr' was built under R version 4.0.3

## Warning: package 'dplyr' was built under R version 4.0.3

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(lmtest)

## Warning: package 'lmtest' was built under R version 4.0.3

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

dat <- foreign::read.spss('D:/\_\_SEMESTER 5/Data-UASADK01-2021.sav',  
 to.data.frame = T)

## re-encoding from UTF-8

head(dat)

## Y X1 X2 X3 X4 X5 X6  
## 1 Miskin 19 Maksimal SMP Perempuan Tidak Bekerja lebih dari 4 Tidak Menabung  
## 2 Miskin 37 Maksimal SMP Perempuan Tidak Bekerja lebih dari 4 Tidak Menabung  
## 3 Miskin 37 Maksimal SMP Perempuan Tidak Bekerja lebih dari 4 Tidak Menabung  
## 4 Miskin 37 Maksimal SMP Perempuan Tidak Bekerja lebih dari 4 Tidak Menabung  
## 5 Miskin 37 Minimal SMA Laki-laki Tidak Bekerja lebih dari 4 Menabung  
## 6 Miskin 60 Minimal SMA Laki-laki Tidak Bekerja Maksimum 4 Tidak Menabung

str(dat)

## 'data.frame': 420 obs. of 7 variables:  
## $ Y : Factor w/ 2 levels "Tidak Miskin",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ X1: num 19 37 37 37 37 60 62 41 44 41 ...  
## $ X2: Factor w/ 2 levels "Minimal SMA",..: 2 2 2 2 1 1 2 2 2 1 ...  
## $ X3: Factor w/ 2 levels "Laki-laki","Perempuan": 2 2 2 2 1 1 2 1 1 1 ...  
## $ X4: Factor w/ 3 levels "Bekerja di Sektor Formal",..: 3 3 3 3 3 3 3 2 2 2 ...  
## $ X5: Factor w/ 2 levels "Maksimum 4","lebih dari 4": 2 2 2 2 2 1 2 2 2 2 ...  
## $ X6: Factor w/ 2 levels "Menabung","Tidak Menabung": 2 2 2 2 1 2 1 2 2 1 ...  
## - attr(\*, "variable.labels")= Named chr [1:7] "Status Kemiskinan Anak" "Umur KRT" "Pendidikan KRT" "Jenis Kelamin KRT" ...  
## ..- attr(\*, "names")= chr [1:7] "Y" "X1" "X2" "X3" ...  
## - attr(\*, "codepage")= int 65001

dat <- dat %>%   
 mutate\_at(-2, ~ fct\_rev(.x))  
  
  
# Model -------------------------------------------------------------------  
m <- glm(Y~., data = dat, family = binomial())  
summary(m)

##   
## Call:  
## glm(formula = Y ~ ., family = binomial(), data = dat)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4568 -0.8420 0.4322 0.7786 2.1967   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.36170 0.77139 -6.951 3.63e-12 \*\*\*  
## X1 0.03958 0.01190 3.326 0.000880 \*\*\*  
## X2Minimal SMA 0.66790 0.24404 2.737 0.006202 \*\*   
## X3Laki-laki 1.72692 0.31429 5.495 3.91e-08 \*\*\*  
## X4Bekerja di Sektor Informal 1.43371 0.32783 4.373 1.22e-05 \*\*\*  
## X4Bekerja di Sektor Formal 1.73215 0.33675 5.144 2.69e-07 \*\*\*  
## X5Maksimum 4 1.45271 0.25469 5.704 1.17e-08 \*\*\*  
## X6Menabung 0.84988 0.24449 3.476 0.000509 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 563.67 on 419 degrees of freedom  
## Residual deviance: 421.10 on 412 degrees of freedom  
## AIC: 437.1  
##   
## Number of Fisher Scoring iterations: 4

exp(coef(m))

## (Intercept) X1   
## 0.004692909 1.040374218   
## X2Minimal SMA X3Laki-laki   
## 1.950138201 5.623318100   
## X4Bekerja di Sektor Informal X4Bekerja di Sektor Formal   
## 4.194215423 5.652806729   
## X5Maksimum 4 X6Menabung   
## 4.274675824 2.339372753

# Pseudo R2 ---------------------------------------------------------------  
DescTools::PseudoR2(m, c('CoxSnell', 'Nagelkerke', 'McFadden'))

## CoxSnell Nagelkerke McFadden   
## 0.2878385 0.3896581 0.2529311

# Hosmer and Lemeshow test ------------------------------------------------  
# Ho : Model fit dengan data  
ResourceSelection::hoslem.test(m$y, m$fitted.values)

##   
## Hosmer and Lemeshow goodness of fit (GOF) test  
##   
## data: m$y, m$fitted.values  
## X-squared = 7.1204, df = 8, p-value = 0.5237

# No C --------------------------------------------------------------------  
m <- glm((Y=='Miskin')~., data = dat, family = binomial())  
summary(m)

##   
## Call:  
## glm(formula = (Y == "Miskin") ~ ., family = binomial(), data = dat)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1967 -0.7786 -0.4322 0.8420 2.4568   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.36170 0.77139 6.951 3.63e-12 \*\*\*  
## X1 -0.03958 0.01190 -3.326 0.000880 \*\*\*  
## X2Minimal SMA -0.66790 0.24404 -2.737 0.006202 \*\*   
## X3Laki-laki -1.72692 0.31429 -5.495 3.91e-08 \*\*\*  
## X4Bekerja di Sektor Informal -1.43371 0.32783 -4.373 1.22e-05 \*\*\*  
## X4Bekerja di Sektor Formal -1.73215 0.33675 -5.144 2.69e-07 \*\*\*  
## X5Maksimum 4 -1.45271 0.25469 -5.704 1.17e-08 \*\*\*  
## X6Menabung -0.84988 0.24449 -3.476 0.000509 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 563.67 on 419 degrees of freedom  
## Residual deviance: 421.10 on 412 degrees of freedom  
## AIC: 437.1  
##   
## Number of Fisher Scoring iterations: 4

x <- sum(coef(m)[-2]\*c(1, 1, 0, 0, 0, 1, 1))  
  
# Minimal Umur  
(log(1) - x)/coef(m)[2]

## X1   
## 60.41392

# Bukti  
newdata <-   
 data.frame(  
 X1 = 60.41392,  
 X2 = 'Minimal SMA',  
 X3 = 'Perempuan',  
 X4 = 'Tidak Bekerja',  
 X5 = 'Maksimum 4',  
 X6 = 'Menabung'  
 )  
newdata

## X1 X2 X3 X4 X5 X6  
## 1 60.41392 Minimal SMA Perempuan Tidak Bekerja Maksimum 4 Menabung

predict(m, newdata, type = 'resp')

## 1   
## 0.5