Assignment 3

Sonali Singh

# Knowledge Mining: Linear Discriminant Analysis

File: Lab\_LDA01.R Theme: Linear Discriminant Analysis Adapted from ISLR Chapter 4 Lab

### Overview

In this assignment, the TEDS\_2016 dataset was used to explore the relationships between various demographic and economic variables with political preferences in Taiwan. The analysis involved running multiple regression models to understand these relationships and the development of a custom function for plotting regression analyses.

### Data Summary

The dataset consists of several variables, including demographic factors such as age, education, and income, and political variables such as party preference and vote for Tsai Ing-wen. A summary of the data revealed:

* **Age** ranges from 20 to 100 years, highlighting a broad demographic spectrum.
* **Education** levels are encoded numerically from 1 (below elementary) to 5 (college and above).
* **Income** also varies widely across respondents, from low to high economic backgrounds.

library(haven)  
TEDS\_2016 <- read\_stata("https://github.com/datageneration/home/blob/master/DataProgramming/data/TEDS\_2016.dta?raw=true")  
  
summary(TEDS\_2016)

## District Sex Age Edu Arear   
## Min. : 201 Min. :1.000 Min. :1.0 Min. :1.000 Min. :1.000   
## 1st Qu.:1401 1st Qu.:1.000 1st Qu.:2.0 1st Qu.:2.000 1st Qu.:1.000   
## Median :6406 Median :1.000 Median :3.0 Median :3.000 Median :3.000   
## Mean :4661 Mean :1.486 Mean :3.3 Mean :3.334 Mean :2.744   
## 3rd Qu.:6604 3rd Qu.:2.000 3rd Qu.:5.0 3rd Qu.:5.000 3rd Qu.:4.000   
## Max. :6806 Max. :2.000 Max. :5.0 Max. :9.000 Max. :6.000   
##   
## Career Career8 Ethnic Party   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. : 1.00   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.: 5.00   
## Median :2.000 Median :4.000 Median :1.000 Median : 7.00   
## Mean :2.683 Mean :3.811 Mean :1.658 Mean :13.02   
## 3rd Qu.:4.000 3rd Qu.:5.000 3rd Qu.:2.000 3rd Qu.:25.00   
## Max. :5.000 Max. :8.000 Max. :9.000 Max. :26.00   
##   
## PartyID Tondu Tondu3 nI2   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. : 1.00   
## 1st Qu.:2.000 1st Qu.:3.000 1st Qu.:2.000 1st Qu.: 1.00   
## Median :2.000 Median :4.000 Median :2.000 Median : 3.00   
## Mean :4.522 Mean :4.127 Mean :2.667 Mean :35.13   
## 3rd Qu.:9.000 3rd Qu.:5.000 3rd Qu.:3.000 3rd Qu.:98.00   
## Max. :9.000 Max. :9.000 Max. :9.000 Max. :98.00   
##   
## votetsai green votetsai\_nm votetsai\_all   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.0000 Median :0.0000 Median :1.0000 Median :1.0000   
## Mean :0.6265 Mean :0.3781 Mean :0.6265 Mean :0.5478   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## NA's :429 NA's :429 NA's :248   
## Independence Unification sq Taiwanese   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :1.0000 Median :1.0000   
## Mean :0.2888 Mean :0.1225 Mean :0.5172 Mean :0.6272   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## edu female whitecollar lowincome   
## Min. :1.000 Min. :0.0000 Min. :0.0000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:4.000   
## Median :3.000 Median :0.0000 Median :1.0000 Median :5.000   
## Mean :3.301 Mean :0.4864 Mean :0.5373 Mean :4.343   
## 3rd Qu.:5.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:5.000   
## Max. :5.000 Max. :1.0000 Max. :1.0000 Max. :5.000   
## NA's :10   
## income income\_nm age KMT   
## Min. : 1.000 Min. : 1.000 Min. : 20.00 Min. :0.0000   
## 1st Qu.: 3.000 1st Qu.: 2.000 1st Qu.: 35.00 1st Qu.:0.0000   
## Median : 5.500 Median : 5.000 Median : 49.00 Median :0.0000   
## Mean : 5.324 Mean : 5.281 Mean : 49.11 Mean :0.2296   
## 3rd Qu.: 7.000 3rd Qu.: 8.000 3rd Qu.: 61.00 3rd Qu.:0.0000   
## Max. :10.000 Max. :10.000 Max. :100.00 Max. :1.0000   
## NA's :330   
## DPP npp noparty pfp   
## Min. :0.0000 Min. :0.00000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.0000 Median :0.00000 Median :0.0000 Median :0.00000   
## Mean :0.3497 Mean :0.02544 Mean :0.3716 Mean :0.01893   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :1.00000   
##   
## South north Minnan\_father Mainland\_father   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :1.0000 Median :0.0000   
## Mean :0.4947 Mean :0.4799 Mean :0.7225 Mean :0.1024   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## Econ\_worse Inequality inequality5 econworse5   
## Min. :0.0000 Min. :0.0000 Min. :1.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:4.000 1st Qu.:3.000   
## Median :1.0000 Median :1.0000 Median :5.000 Median :4.000   
## Mean :0.5544 Mean :0.9355 Mean :4.495 Mean :3.644   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:5.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :1.0000 Max. :5.000 Max. :5.000   
##   
## Govt\_for\_public pubwelf5 Govt\_dont\_care highincome   
## Min. :0.0000 Min. :1.000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :3.000 Median :0.0000 Median :1.0000   
## Mean :0.4249 Mean :2.877 Mean :0.4988 Mean :0.5765   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :5.000 Max. :1.0000 Max. :1.0000   
## NA's :330   
## votekmt votekmt\_nm Blue Green No\_Party  
## Min. :0.0000 Min. :0.0000 Min. :0 Min. :0 Min. :0   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0 1st Qu.:0 1st Qu.:0   
## Median :0.0000 Median :0.0000 Median :0 Median :0 Median :0   
## Mean :0.2053 Mean :0.2752 Mean :0 Mean :0 Mean :0   
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0   
## Max. :1.0000 Max. :1.0000 Max. :0 Max. :0 Max. :0   
## NA's :429   
## voteblue voteblue\_nm votedpp\_1 votekmt\_1   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :1.0000 Median :0.0000   
## Mean :0.2787 Mean :0.3735 Mean :0.5256 Mean :0.2309   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## NA's :429 NA's :187 NA's :187

str(TEDS\_2016)

## tibble [1,690 × 54] (S3: tbl\_df/tbl/data.frame)  
## $ District : dbl+lbl [1:1690] 201, 201, 201, 201, 201, 201, 201, 201, 201, 201, 201...  
## ..@ label : chr "District"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:73] 201 401 501 502 701 702 703 704 801 802 ...  
## .. ..- attr(\*, "names")= chr [1:73] "Yi Lan County Single District" "Hsinchu County Single District" "Miaoli County 1st District" "Miaoli County 2nd District" ...  
## $ Sex : dbl+lbl [1:1690] 2, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1,...  
## ..@ label : chr "Sex"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Male" "Female"  
## $ Age : dbl+lbl [1:1690] 4, 2, 5, 4, 5, 5, 5, 4, 5, 4, 5, 1, 5, 3, 4, 5, 4, 5,...  
## ..@ label : chr "Age"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:5] 1 2 3 4 5  
## .. ..- attr(\*, "names")= chr [1:5] "20-29" "30-39" "40-49" "50-59" ...  
## $ Edu : dbl+lbl [1:1690] 4, 5, 5, 2, 1, 2, 1, 5, 1, 1, 1, 2, 1, 5, 5, 1, 3, 4,...  
## ..@ label : chr "Education"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:6] 1 2 3 4 5 9  
## .. ..- attr(\*, "names")= chr [1:6] "Below elementary school" "Junior high school" "Senior high school" "College" ...  
## $ Arear : dbl+lbl [1:1690] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Area"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:6] 1 2 3 4 5 6  
## .. ..- attr(\*, "names")= chr [1:6] "Taipei, New Taipei, Keelung and Yi Lan" "Taoyuan, Hsinchu and Miaoli" "Taichung, Changhua and Nantou" "Yunlin, Chiayi and Tainan" ...  
## $ Career : dbl+lbl [1:1690] 1, 2, 1, 4, 3, 2, 4, 1, 4, 3, 3, 5, 5, 4, 1, 5, 2, 2,...  
## ..@ label : chr "Occupations5"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:5] 1 2 3 4 5  
## .. ..- attr(\*, "names")= chr [1:5] "Hight-class WHITE COLLAR" "Low-class WHITE COLLAR" "FARMER" "WORKER" ...  
## $ Career8 : dbl+lbl [1:1690] 1, 3, 1, 4, 5, 7, 4, 2, 4, 5, 5, 7, 7, 7, 2, 7, 3, 1,...  
## ..@ label : chr "Occupation8"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:8] 1 2 3 4 5 6 7 8  
## .. ..- attr(\*, "names")= chr [1:8] "Civil servants" "Managers and Professionals (priv.)" "CLERKS (priv.)" "Labor (priv.)" ...  
## $ Ethnic : dbl+lbl [1:1690] 1, 2, 2, 1, 9, 1, 2, 1, 1, 2, 1, 1, 2, 1, 2, 9, 2, 2,...  
## ..@ label : chr "Ethnic"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:4] 1 2 3 9  
## .. ..- attr(\*, "names")= chr [1:4] "Taiwanese" "Both" "Chinese" "Noresponse"  
## $ Party : dbl+lbl [1:1690] 25, 25, 3, 25, 25, 6, 25, 24, 25, 25, 6, 5, 25, ...  
## ..@ label : chr "Party Preference"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:26] 1 2 3 4 5 6 7 8 9 10 ...  
## .. ..- attr(\*, "names")= chr [1:26] "Strongly support KMT" "Somewhat support KMT" "Lean to KMT" "Somewhat lean to KMT" ...  
## $ PartyID : dbl+lbl [1:1690] 9, 9, 1, 9, 9, 2, 9, 6, 9, 9, 2, 2, 9, 1, 1, 9, 9, 9,...  
## ..@ label : chr "Party Identification"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:7] 1 2 3 4 5 6 9  
## .. ..- attr(\*, "names")= chr [1:7] "KMT" "DPP" "NP" "PFP" ...  
## $ Tondu : dbl+lbl [1:1690] 3, 5, 3, 5, 9, 4, 9, 6, 9, 9, 5, 5, 9, 5, 4, 9, 9, 4,...  
## ..@ label : chr "Position on unification and independence"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:7] 1 2 3 4 5 6 9  
## .. ..- attr(\*, "names")= chr [1:7] "Immediate unification" "Maintain the status quo,move toward unification" "Maintain the status quo, decide either unification or independence" "Maintain the status quo forever" ...  
## $ Tondu3 : dbl+lbl [1:1690] 2, 3, 2, 3, 9, 2, 9, 3, 9, 9, 3, 3, 9, 3, 2, 9, 9, 2,...  
## ..@ label : chr "3 categories of TONDU"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:4] 1 2 3 9  
## .. ..- attr(\*, "names")= chr [1:4] "Unification" "Maintain the status quo" "Independence" "Nonresponse"  
## $ nI2 : dbl+lbl [1:1690] 3, 98, 98, 3, 98, 98, 98, 3, 98, 1, 2, 98, 98, ...  
## ..@ label : chr "Who is the current the premier of our country?"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:5] 1 2 3 95 98  
## .. ..- attr(\*, "names")= chr [1:5] "Correct" "Incorrect" "I know but can't remember the name" "Refuse to answer" ...  
## $ votetsai : num [1:1690] NA 1 0 NA NA 1 1 1 1 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ green : num [1:1690] 0 0 0 0 0 1 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votetsai\_nm : num [1:1690] NA 1 0 NA NA 1 1 1 1 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votetsai\_all : num [1:1690] 0 1 0 0 0 1 1 1 1 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Independence : num [1:1690] 0 1 0 1 0 0 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Unification : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ sq : num [1:1690] 1 0 1 0 0 1 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Taiwanese : num [1:1690] 1 0 0 1 0 1 0 1 1 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ edu : num [1:1690] 4 5 5 2 1 2 1 5 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ female : num [1:1690] 1 1 0 0 1 1 0 1 1 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ whitecollar : num [1:1690] 1 1 1 0 0 1 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ lowincome : num [1:1690] 4 4 5 4 3 5 2 5 5 5 ...  
## ..- attr(\*, "label")= chr "How serious do you think low income of salaryman?"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ income : num [1:1690] 8 7 8 5 5.5 9 1 10 2 5.5 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ income\_nm : num [1:1690] 8 7 8 5 NA 9 1 10 2 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ age : num [1:1690] 59 39 63 55 76 64 75 54 64 59 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ KMT : num [1:1690] 0 0 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ DPP : num [1:1690] 0 0 0 0 0 1 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ npp : num [1:1690] 0 0 0 0 0 0 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ noparty : num [1:1690] 1 1 0 1 1 0 1 0 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ pfp : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ South : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ north : num [1:1690] 1 1 1 1 1 1 1 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Minnan\_father : num [1:1690] 1 1 1 1 1 1 1 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Mainland\_father: num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Econ\_worse : num [1:1690] 0 0 1 1 0 1 1 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Inequality : num [1:1690] 1 1 1 1 0 1 0 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ inequality5 : num [1:1690] 4 5 5 5 3 5 3 5 5 5 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ econworse5 : num [1:1690] 3 3 4 5 3 4 4 5 5 5 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Govt\_for\_public: num [1:1690] 1 1 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ pubwelf5 : num [1:1690] 5 5 4 1 3 2 2 1 3 2 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Govt\_dont\_care : num [1:1690] 0 0 1 1 0 1 1 1 0 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ highincome : num [1:1690] 1 1 1 1 NA 1 0 1 0 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votekmt : num [1:1690] 0 0 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votekmt\_nm : num [1:1690] NA 0 1 NA NA 0 0 0 0 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Blue : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Green : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ No\_Party : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ voteblue : num [1:1690] 0 0 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ voteblue\_nm : num [1:1690] NA 0 1 NA NA 0 0 0 0 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votedpp\_1 : num [1:1690] NA 1 0 NA NA 1 1 1 1 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votekmt\_1 : num [1:1690] NA 0 1 NA NA 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"

regplot <- function(x, y) {  
 fit <- lm(y ~ x)  
 plot(x, y, main = paste("Regression Plot of", deparse(substitute(y)), "on", deparse(substitute(x))))  
 abline(fit, col = "red")  
}  
  
# Ensure the variables are appropriate for regression; typically needs to be continuous  
# Let's look at the structure of these variables first  
str(TEDS\_2016$age)

## num [1:1690] 59 39 63 55 76 64 75 54 64 59 ...  
## - attr(\*, "format.stata")= chr "%9.0g"

str(TEDS\_2016$edu)

## num [1:1690] 4 5 5 2 1 2 1 5 1 1 ...  
## - attr(\*, "format.stata")= chr "%9.0g"

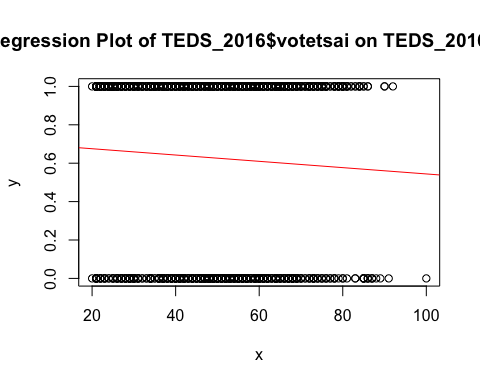
str(TEDS\_2016$income)

## num [1:1690] 8 7 8 5 5.5 9 1 10 2 5.5 ...  
## - attr(\*, "format.stata")= chr "%9.0g"

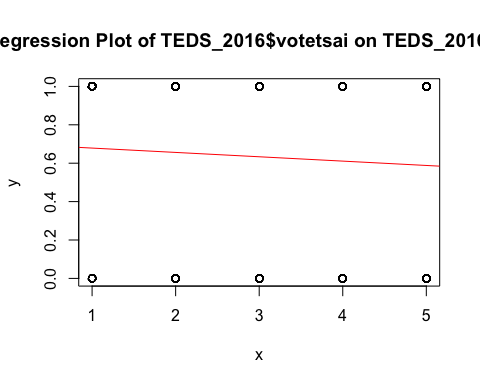
# Convert factors to numeric if they are not already  
TEDS\_2016$age <- as.numeric(TEDS\_2016$age)  
TEDS\_2016$edu <- as.numeric(TEDS\_2016$edu)  
TEDS\_2016$income <- as.numeric(TEDS\_2016$income)  
  
# Assuming `votetsai` is the dependent variable, check its structure  
str(TEDS\_2016$votetsai)

## num [1:1690] NA 1 0 NA NA 1 1 1 1 NA ...  
## - attr(\*, "format.stata")= chr "%9.0g"

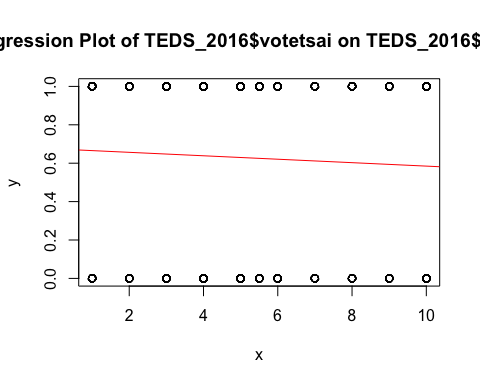
# Run regplot for Age, Education, and Income against votetsai  
regplot(TEDS\_2016$age, TEDS\_2016$votetsai)



regplot(TEDS\_2016$edu, TEDS\_2016$votetsai)



regplot(TEDS\_2016$income, TEDS\_2016$votetsai)



# If `votetsai` is binary, consider logistic regression instead  
fit\_logit <- glm(votetsai ~ age + edu + income, data = TEDS\_2016, family = binomial)  
summary(fit\_logit)

##   
## Call:  
## glm(formula = votetsai ~ age + edu + income, family = binomial,   
## data = TEDS\_2016)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.508583 0.379569 6.609 3.87e-11 \*\*\*  
## age -0.021731 0.004656 -4.668 3.05e-06 \*\*\*  
## edu -0.238448 0.054069 -4.410 1.03e-05 \*\*\*  
## income -0.020470 0.022460 -0.911 0.362   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1661.8 on 1256 degrees of freedom  
## Residual deviance: 1632.0 on 1253 degrees of freedom  
## (433 observations deleted due to missingness)  
## AIC: 1640  
##   
## Number of Fisher Scoring iterations: 4

### Regression Plot Function

A function named **regplot** was developed to facilitate the plotting of regression lines for various explanatory variables against a chosen dependent variable. This function integrates the **lm**, **plot**, and **abline** functions to create a clear visual representation of the relationship between two variables.

### Regression Analysis

Multiple regression plots were generated using the **regplot** function to examine the effects of age, education, and income on the likelihood of voting for Tsai Ing-wen:

* **Age:** The regression analysis suggested a negative relationship, indicating that younger individuals were more likely to vote for Tsai Ing-wen.
* **Education:** Higher education levels appeared to decrease the likelihood of voting for Tsai, possibly reflecting different political leanings across educational groups.
* **Income:** The relationship was less clear, with no significant effect observed in the preliminary regression analysis.

### Challenges Encountered

The primary challenge was dealing with the binary nature of the dependent variable (**votetsai**), which indicates whether respondents voted for Tsai Ing-wen (1) or not (0). Standard linear regression is not well-suited for binary outcomes, which led to the consideration of logistic regression.

### Logistic Regression

Given the binary outcome of the dependent variable, a logistic regression model was more appropriate. This model showed:

* **Age** and **education** significantly affected the probability of voting for Tsai, with both showing negative coefficients, suggesting that younger, less educated individuals were more likely to vote for Tsai.
* **Income** did not significantly influence voting behavior, indicating economic status might not be a strong predictor in this electoral context.

### Conclusion

The exploratory data analysis provided valuable insights into the factors influencing voting behavior in Taiwan. The custom **regplot** function proved useful for visualizing these relationships, although the analysis highlighted the importance of choosing appropriate models for the nature of the data. Future analyses could expand on these findings by incorporating additional variables or using different subsets of the data to further uncover the dynamics of voter preferences.

### Recommendations

* Further exploration with more complex models like multinomial logistic regression could be considered if the dependent variable categories expand beyond binary choices.
* Additional variables such as geographical location or more detailed income brackets could be included to enhance the model’s predictive power and accuracy.

This analysis serves as a foundational step in understanding the interplay between socio-economic factors and political preferences, offering a methodological framework for future studies in similar contexts.