# Assignment10

## #question1

library(tidyverse)

my\_defaults <- read.csv("defaults.csv")

set.seed(17053777)

my\_defaults <- my\_defaults %>%

mutate(student = factor(student)) %>%

mutate(default = factor(default)) %>%

sample\_n(600)

## #question2

library(farver)

my\_defaults %>%

ggplot(aes(x = balance, y = default, colour = student)) +

geom\_jitter(alpha = 0.6, size = 2, height = 0.1) +

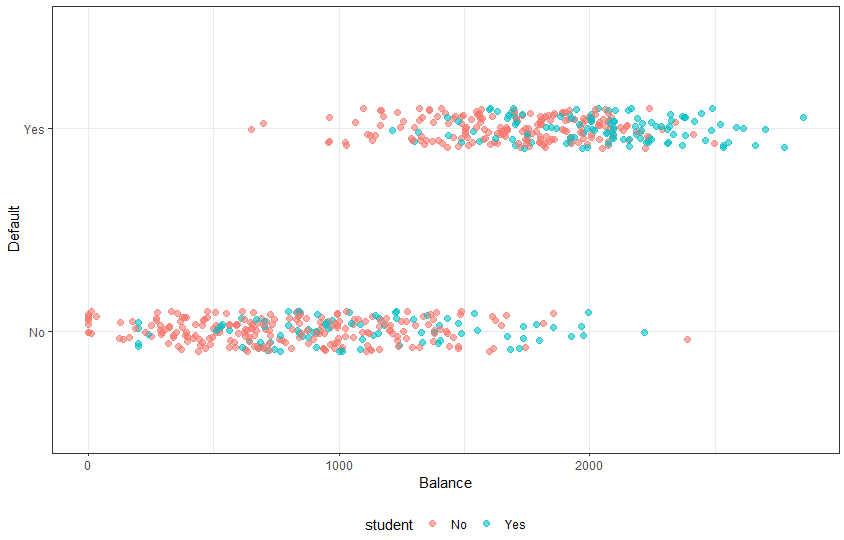
labs(x = "Balance", y = "Default") +

theme\_bw() +

theme(legend.position = "bottom")

# Yes, we can predict a credit card holder will default using their credit card balance.

# It does not depend on whether they are student or not since there is no obvious separation between students and non-students.



## #Question3

m1 <- glm(default ~ balance +

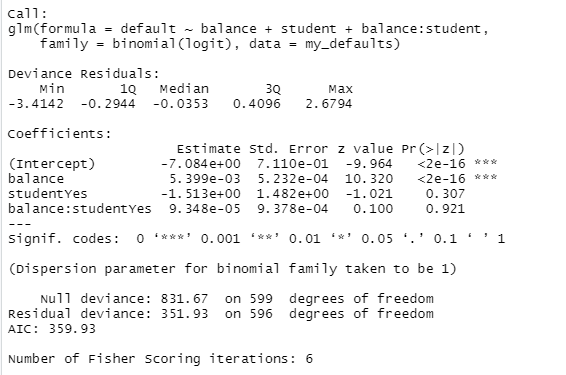
student +

balance:student,

family = binomial(logit),

data = my\_defaults )

summary(m1)



## #question4

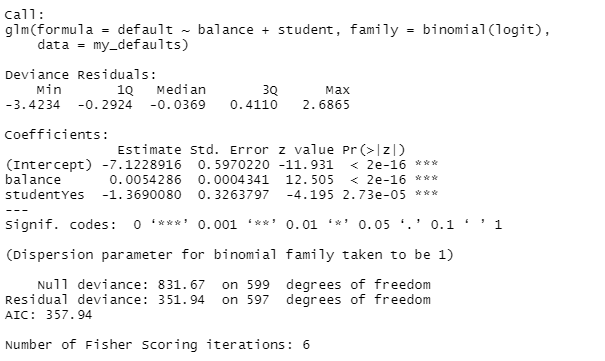
m2 <- glm(default ~ balance +

student,

family = binomial(logit),

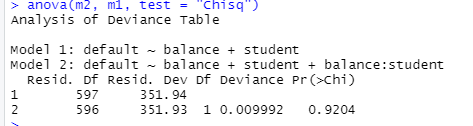
data = my\_defaults)

summary(m2)



## #question5

anova(m2, m1, test = "Chisq")



# p-value from Chi square test is 0.9204, p value is greater than 0.05, so we don't reject.

# null hypothesis: there is no interaction between balance and student.

# As we don’t reject H0, we can say the interaction between balance and student is not statistically significant.

## #question6

# From the summary of m2, log(P/(1-P)) = -7.12 + 0.0054 \* balance – 1.37 \* student

# student: 0 for No, 1 for Yes

# balance: 0 for No, 1 for Yes

# i)

# log(P/(1-P)) = -7.12

# odds = p/(1-P) = e ^ (-7.12)

# p = 1/(1 + e^ (-(-7.12))) = 0.000808

# If that is non-student with balance of 0, there is a really small chance to default.

# ii)

# 0 for No, 1 for Yes

# log(P/(1-P)) = -7.12 - 1.37 \* 1 = -8.49

# odds = p/(1-P) = e ^ (-8.49)

# p = 1/(1 + e^ (-(-8.49))) = 0.000205471

# If that is a student with balance of 0, there is also a really small chance to default.

# iii)

# For a $1 increase in balance, the odds of default are multiples by exp(0.0054286)

## # question7

new\_data <- tibble(balance = 1500, student = c("No", "Yes"))

predict(m2, new\_data, type = "response")



# For non-student, there will be 0.7349841 probability to default if with balance $1500.

# For student, there will be 0.4136374 probability to default if with balance $1500.

# There is a higher probability for non-student to default with a balance of $1500.

## #question8

# Apparent Error Ratio = (False Positives + False negatives) / total number

# Apparent Error Ratio = (42 + 33) / 600 = 0.125

# save coefficient values

b0\_N <- coef(m2)[[1]] # intercept, not student

b1\_N <- coef(m2)[[2]] # logit slope, not student

# you need to create variables for b0\_Y and b1\_Y

# create b0\_Y the intercept for students

# create b1\_Y the logit slope for students

# plot the data and curves # balances data for plotting

plot\_balances <- tibble(balance = seq(0, 3000, 500))

ggplot(my\_defaults, aes(

x = balance,

y = ifelse(default == "No", -0.1, 1.1), # plot observed points below 0 and above 1

colour = student )) + geom\_jitter(alpha = 0.8, height = 0.1) +

# plot curve for not students (assumes you have variables for b0\_N, b1\_N)

geom\_function( data = plot\_balances, aes(x = balance), inherit.aes = FALSE,# does not use parent ggplot's aesthetics

fun = function(x) plogis(b0\_N + b1\_N \* x), # plogis does expit conversion

size = 1, colour = "#F8766D" # emulate ggplot default colours :)

) + geom\_function(

# plot curve for students (assumes you have variables for b0\_Y, b1\_Y)

data = plot\_balances, aes(x = balance),

inherit.aes = FALSE,

fun = function(x) plogis(b0\_Y + b1\_Y \* x),

size = 1, colour = "#00BFC4" # emulate ggplot default colours :)

) + # tidy up R's default y-axis

scale\_y\_continuous(expand = c(0, 0), limits = c(-0.2, 1.2), breaks = c(0, 1)) +

labs(x = "balance", y = "y = 1 if default, 0 otherwise\n and modeled probabilities") +

theme\_bw() +

theme(legend.position = "bottom")

