

Final Report

due November 16, 2021 by 11:59 PM

Danielle Mensah, Haby Sow, Colin Lee

11/12/2021

```
library(tidyverse)

## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.5       v dplyr 1.0.7
## v tidyr 1.1.4        v stringr 1.4.0
## v readr 2.0.2        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(tidymodels)

## Registered S3 method overwritten by 'tune':
##   method                from
##   required_pkgs.model_spec parsnip

## -- Attaching packages ----- tidymodels 0.1.4 --

## v broom      0.7.9      v rsample      0.1.0
## v dials      0.0.10     v tune         0.1.6
## v infer      1.0.0      v workflows    0.2.4
## v modeldata  0.1.1      v workflowsets 0.1.0
## v parsnip    0.1.7      v yardstick    0.0.8
## v recipes    0.1.17

## -- Conflicts ----- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Learn how to get started at https://www.tidymodels.org/start/

setwd("/home/guest/R/project01")
student <- readr::read_csv("data/student-mat.csv")

## Rows: 395 Columns: 33

## -- Column specification -----
```

```

## Delimiter: ","
## chr (17): school, sex, address, famsize, Pstatus, Mjob, Fjob, reason, guardi...
## dbl (16): age, Medu, Fedu, traveltime, studytime, failures, famrel, freetime...

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
student_binger <- student %>%
  mutate(binger = ifelse(sex == "F", ifelse(Dalc >= 3,1,0), ifelse(Dalc >= 4, 1, 0)))

student_binger$binger=factor(student_binger$binger,levels=c(1,0),labels=c("Yes","No"))
student_binger$binger=relevel(student_binger$binger, ref = "No")

student_logit <- student_binger %>%
  mutate(urban = ifelse(address == "U", 1, 0)) %>%
  mutate(famlarge = ifelse(famsize == "GT3", 1, 0)) %>%
  mutate(parents_together = ifelse(Pstatus == "T", 1, 0)) %>%
  mutate(mother_secondary = ifelse(Medu >= 3, 1, 0)) %>%
  mutate(father_secondary = ifelse(Fedu >= 3, 1, 0)) %>%
  mutate(school_support = ifelse(schoolsup == "yes", 1, 0)) %>%
  mutate(family_support = ifelse(famsup == "yes", 1, 0)) %>%
  mutate(extra_tutoring = ifelse(paid == "yes", 1, 0)) %>%
  mutate(alcoholic = ifelse(binger == "Yes", 1, 0))

student_logit_fit <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(binger ~ urban + famlarge + parents_together + mother_secondary + father_secondary + school_support)

tidy(student_logit_fit, conf.int=TRUE, exponentiate = TRUE)

## # A tibble: 9 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)         0.0447    0.913    -3.41    0.000660 0.00630    0.235
## 2 urban               0.930     0.529    -0.137   0.891     0.351     2.92
## 3 famlarge            0.764     0.468    -0.576   0.565     0.312     2.00
## 4 parents_together    0.771     0.673    -0.387   0.699     0.232     3.53
## 5 mother_secondary    1.26     0.546     0.426   0.670     0.436     3.77
## 6 father_secondary    0.952     0.511    -0.0965  0.923     0.353     2.65
## 7 school_support      1.14     0.660     0.198   0.843     0.254     3.68
## 8 family_support      0.800     0.486    -0.459   0.646     0.315     2.16
## 9 extra_tutoring      3.84     0.511     2.63    0.00847  1.48     11.3

#Clearing Missing Data
missingval <- is.na(student)
head(missingval)

##   school  sex  age address famsize Pstatus  Medu  Fedu  Mjob  Fjob reason
## [1,] FALSE FALSE FALSE  FALSE  FALSE  FALSE FALSE FALSE FALSE FALSE FALSE
## [2,] FALSE FALSE FALSE  FALSE  FALSE  FALSE FALSE FALSE FALSE FALSE FALSE
## [3,] FALSE FALSE FALSE  FALSE  FALSE  FALSE FALSE FALSE FALSE FALSE FALSE
## [4,] FALSE FALSE FALSE  FALSE  FALSE  FALSE FALSE FALSE FALSE FALSE FALSE
## [5,] FALSE FALSE FALSE  FALSE  FALSE  FALSE FALSE FALSE FALSE FALSE FALSE
## [6,] FALSE FALSE FALSE  FALSE  FALSE  FALSE FALSE FALSE FALSE FALSE FALSE
##   guardian traveltime studytime failures schoolsup famsup  paid activities

```

```

## [1,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [2,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [3,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [4,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [5,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##      nursery higher internet romantic famrel freetime goout Dalc Walc health
## [1,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [2,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [3,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [4,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [5,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##      absences      G1      G2      G3
## [1,] FALSE FALSE FALSE FALSE
## [2,] FALSE FALSE FALSE FALSE
## [3,] FALSE FALSE FALSE FALSE
## [4,] FALSE FALSE FALSE FALSE
## [5,] FALSE FALSE FALSE FALSE
## [6,] FALSE FALSE FALSE FALSE

```

As you can see from this quick check. There are no missing values in our data. Therefore we can move on with further analysis and no clearing of variables needs to be done. I put only the head of the data because it was too long to visually see the whole thing however it is all false.

#Data Wrangling There are two big questions that we want answered with this data set: whether a students average alcohol consumption is correlated with their family circumstances and whether alcohol consumption has an effect on student life. Lets first look at some geographgics of our students