**Attempt coding Ogochukwu’s project by Nicole****Antunes Rezende**

Research question:

Is there any significant association between total income and sex, age at immigration, and level of education of Canadian Immigrants from 2000 to 2015 who immigrated at the age of 20 years and older?

Dataset: Census 2016.csv

1. Install necessary packages and load the libraries

Nicole’s comment: The first instruction was “Install necessary packages and load the libraries”. It was impossible to imagine which are the necessary packages for this project without previous instructions.

2. Clean up data by:

a. Loading the CSV file

install.packages(‘readr’) library (readr)

read\_csv(‘pumf-98M0001-E-2016-individuals\_F1.csv’)

census <- read\_csv(‘pumf-98M0001-E-2016-individuals\_F1.csv’)

b. Select only TotInc, Sex, AGEIMM, HDGREE, YRIMM (variables) and discard others

install.packages(“dplyr”) library(dplyr)

selected\_variables <- census |> select(TotInc, Sex, AGEIMM, HDGREE, YRIMM)

print(selected\_variables)

c. Filter AGEIMM and select range 5 to 13 [meaning 20 years and above]

selected\_variables <- selected\_variables |> filter (AGEIMM >=5 & AGEIMM <=13)

print(selected\_variables)

d. Remove cases with TotInc less than 1000 and those labelled 88888888 and 99999999

selected\_variables <- selected\_variables |> filter (!(TotInc <=1000 | TotInc==88888888 | TotInc==99999999))

print(selected\_variables)

e. Remove cases with HDGREE = 88 or 99

selected\_variables <- selected\_variables |> filter (!(HDGREE==88 | HDGREE == 99))

print(selected\_variables)

f. Convert all variables in the dataset to numeric

Nicole’s comment: Not clear for me this instruction as using the function “glimpse”, I see that the variables are dbl (double), which is one of the main numeric types in R.

glimpse (selected\_variables)

selected\_variables <- selected\_variables |> mutate(across(everything(), as.numeric))

glimpse (selected\_variables)

Nicole’s comment: I used the function “mutate”, but anything changed as all the variables were already dbl. So, I’m not sure what was the result wanted with this instruction.

g. Re-code Sex as Female =1, Male = 2

Nicole’s comment: I don’t understand this instruction “Re-code Sex as Female =1, Male = 2”, as in the data set Female is already 1 and male is already 2.

h. Creating HDGREE.R variable such that 1 to 8 = below bachelor, 9 = bachelor, and 10 to 13 = above bachelor

table (selected\_variablesHDGREE

library(dplyr)

selected\_variables <- selected\_variables |> mutate(HDGREE.R = case\_when ( HDGREE >= 1 & HDGREE <= 8 ~ 1, HDGREE ==9 ~ 2, HDGREE >= 10 & HDGREE <= 13 ~ 3 ))

selected\_variablesHDGREE.R)

selected\_variablesHDGREE.R, levels = c(1, 2 ,3), labels = c(“below bachelor”, “bachelor”, “above bachelor” ))

selected\_variablesHDGREE.R)

3. Run descriptive statistics to create frequency and percentage summary table for Sex and Education levels.

summarytable1 <- table(selected\_variablesHDGREE.R) print(summarytable1)

table (selected\_variablesSex)

summarytable2 <- prop.table(summarytable1) \*100 summarytable2[] <- sprintf(“%.2f%%”, summarytable2) print(summarytable2)

4. Run a multivariate regression analysis with TotInc as the dependent variable and Sex, age at immigration, and education qualification (HDGREE.R) as independent variables.

model <- lm(TotInc ~ Sex + AGEIMM + HDGREE.R, data = selected\_variables)

summary (model)

Nicole’s comment: We are not answering all aspects of the research question, as we are not selecting the period (from 2010 and 2015). Research question:

“Is there any significant association between total income and sex, age at immigration, and level of education of Canadian Immigrants from 2000 to 2015 who immigrated at the age of 20 years and older?”

5. Write the 2 or 3 tabular output in any file format.

Nicole’s comment: I was not able to save the outputs in word, pdf or htlm. I have tried with the following code, but it did not work.

CODE not working

Output in Word:

install.packages(c(“officer”, “flextable”))

library(officer)

library(flextable)

doc <- read\_docx()

doc <- doc |>

body\_add\_par(“Summary Table 1”, style = “heading 1”) |>

body\_add\_flextable(flextable(as.data.frame(summarytable1))) |>

body\_add\_par(“Summary Table 2 (Proportions)”, style = “heading 1”) |>

body\_add\_flextable(flextable(as.data.frame(summarytable2)))

print(doc, target = “summary\_tables.docx”)

Output in htlm using rmardown

install.packages(“rmarkdown”)

install.packages(“knitr”)

library(knitr)

summarytable <- kable(summarytable1 ,format = “html”, caption = “Frequency table”)

kable(summarytable1 ,format = “html”, caption = “Frequency table”)