MSc Dissertation

220225605

2023-08-31

# MSc Dissertation

## Does ethnicity and income moderate the relationship between the consumption of saturated fats and free sugars and BMI?

### Loading the required libraries

library(here)

## here() starts at C:/Users/solom/Desktop/Final Project MSc/MSc Dissertation

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(naniar)

### Importing required datasets (Waves 1 to 4)

# opening person level dietary records for waves 1 to 4  
PLDR14 <- read\_csv(here("data","PersonLevel1-4.csv"))

## Rows: 6828 Columns: 379  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Country  
## dbl (372): seriali, Age, Sex, SurveyYear, Ndays, TotalEMJ, FoodEMJ, EnergykJ...  
## lgl (6): bloThiamlrni, bloNiacElrni, bloVitB6lrni, bloThiamlrniplussupps, ...  
##   
## 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.

# retaining required columns from the data-set above   
PLDR14 <- PLDR14 %>% select(seriali, Ndays, Saturatedfattyacidsg, FreeSugarsg, Energykcal)   
  
# opening individual level data for waves 1 to 4   
Indiv14 <- read\_csv(here("data", "Indiv1-4.csv"))

## Rows: 6828 Columns: 2147  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (54): ActNo, ActNoy2, StadNo, SclNo, LatitudeS1, LatitudeS2, LatitudeS...  
## dbl (2087): Outcome, PGrid, DMHSize, NumAdult, NumCh118, Num19, NumCh04, Num...  
## lgl (6): LatitudeW6, LatitudeW7, LatitudeW8, LatitudeW9, LatitudeW10, Lat...  
##   
## 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.

# retaining required columns from the data-set above   
Indiv14 <- Indiv14 %>% select(Sex, age, eqvinc, EthGrG, qual7, cigst2, NumCh118, seriali, bmival, CutDown, DMHSize)  
  
# renaming a column so that the datasets (PLDR14 & Indiv14) can be merged   
Indiv14 <- Indiv14 %>% rename(EthGrpGB = EthGrG)  
  
# merging the two datasets together  
Data14 <- merge(x = Indiv14, y = PLDR14, by = "seriali", all.x = TRUE)

### Importing required datasets (Waves 5 to 6)

# opening person level dietary records for waves 5 to 6  
PLDR56 <- read\_csv(here("data","PersonLevel5-6.csv"))

## Rows: 2546 Columns: 381  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Country  
## dbl (374): seriali, Age, Sex, Surveyyear, Ndays, TotalEMJ, FoodEMJ, EnergykJ...  
## lgl (6): bloThiamlrni, bloNiacElrni, bloVitB6lrni, bloThiamlrniplussupps, ...  
##   
## 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.

# retaining required columns from the data-set above   
PLDR56 <- PLDR56 %>% select(seriali, Ndays, Saturatedfattyacidsg, FreeSugarsg, Energykcal)  
  
# opening individual level data for waves 5 to 6   
Indiv56 <- read\_csv(here("data", "Indiv5-6.csv"))

## Rows: 2546 Columns: 1814  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (41): StadNo, SclNo, LatSun1, LatSun2, LatSun3, LatSun4, LatSun5, LatS...  
## dbl (1771): seriali, serialh, Outcome, PGrid, DMHSize, NumAdult, NumCh118, N...  
## lgl (2): LatSun9, latsun10  
##   
## 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.

# retaining required columns from the data-set above   
Indiv56 <- Indiv56 %>% select(Sex, age, eqvinc, EthGrpGB, qual7, cigst2, NumCh118, seriali, bmival, CutDown, DMHSize)  
  
# merging the two datasets together  
Data56 <- merge(x = Indiv56, y = PLDR56, by = "seriali", all.x = TRUE)

### Merging data-sets from 2008/2009 - 2013/2014

FinalData <- rbind(Data14, Data56)

### Removing respondents below the age of 18

FinalData <- FinalData[FinalData$age > 17,]

### Dealing with missing values

# removing rows with the following values: -1 (not applicable), -4 (not applicable to survey year), -8 (Don't know/ Can't say), -9 (Refusal)  
FinalData <- FinalData %>% replace\_with\_na(replace = list(bmival = c(-1, -4, -8, -9), Saturatedfattyacidsg = c(-1,-4, -8, -9), FreeSugarsg = c(-1, -4, -8, -9), eqvinc = c(-1,-4, -8, -9), cigst2 = c(-1,-4, -8, -9), NumCh118 = c(-1,-4, -8, -9), qual7 = c(-1,-4, -8, -9), Energykcal = c(-1,-4, -8, -9), DMHSize = c(-1,-4, -8, -9), CutDown = c(-1,-4, -8, -9), EthGrpGB = c(-1, -4, -8, -9)))  
# drop rows with missing values   
FinalData <- FinalData %>% drop\_na()

### Recoding the ethnicity column

# white  
FinalData$EthGrpGB[FinalData$EthGrpGB == 2] <- 1  
FinalData$EthGrpGB[FinalData$EthGrpGB == 3] <- 1  
FinalData$EthGrpGB[FinalData$EthGrpGB == 4] <- 1  
# south asian   
FinalData$EthGrpGB[FinalData$EthGrpGB == 9] <- 2  
FinalData$EthGrpGB[FinalData$EthGrpGB == 10] <- 2  
FinalData$EthGrpGB[FinalData$EthGrpGB == 11] <- 2  
# black   
FinalData$EthGrpGB[FinalData$EthGrpGB == 13] <- 3  
FinalData$EthGrpGB[FinalData$EthGrpGB == 14] <- 3  
FinalData$EthGrpGB[FinalData$EthGrpGB == 15] <- 3  
  
# chinese   
FinalData$EthGrpGB[FinalData$EthGrpGB == 16] <- 4  
  
# arab   
FinalData$EthGrpGB[FinalData$EthGrpGB == 17] <- 5  
  
# dropping rows for mixed and 'any other' ethnicity groups   
FinalData <- FinalData %>% replace\_with\_na(replace = list(EthGrpGB = c(5, 6, 7, 8, 12, 18)))  
FinalData <- FinalData %>% drop\_na()

### Labelling factor variables

# converting numeric variables to factor variables   
# sex   
FinalData$Sex <- cut(  
 FinalData$Sex, 2, labels = c(  
 'Male',   
 'Female'))  
   
# number of children (converting a continuous variable to a binary, categorical variable)  
FinalData$NumCh118[FinalData$NumCh118 > 0] <- "Have Children"  
FinalData$NumCh118[FinalData$NumCh118 == 0] <- "Do not have Children"  
FinalData$NumCh118 <- as.factor(FinalData$NumCh118)  
  
# cigarette smoking status (converting a polytomous categorical variable to a binary, categorical variable)  
FinalData$cigst2[FinalData$cigst2 == 5] <- "Non-smoker"  
FinalData$cigst2[FinalData$cigst2 < 5] <- "Smoker"  
FinalData$cigst2 <- as.factor(FinalData$cigst2)  
  
# educational qualifications (converting a polytomous categorical variable to a binary, categorical variable)  
FinalData$qual7[FinalData$qual7 <= 2] <- "Higher education and above"  
FinalData$qual7[FinalData$qual7 == 3] <- "A level or equivalent"  
FinalData$qual7[FinalData$qual7 == 4] <- "GCSE or equivalent"  
FinalData$qual7[FinalData$qual7 == 5] <- "GCSE or equivalent"  
FinalData$qual7[FinalData$qual7 == 6] <- "Foreign or other qualifications"  
FinalData$qual7[FinalData$qual7 == 7] <- "No qualifications"  
FinalData$qual7[FinalData$qual7 == 8] <- "Still in FT education"  
FinalData$qual7 <- as.factor(FinalData$qual7)

### Categorise the outcome variable bmival for the logistic regression model

# using a single bmi cutoff point to categorise obesity  
FinalData <- FinalData %>% mutate(bmi\_uni = case\_when(bmival >= 30 ~ "Obese", bmival < 30 ~ "Not Obese"))  
  
# using multiple bmi cutoffs to categorise obesity on the basis of ethnicity   
FinalData <- FinalData %>% mutate(bmi\_eth = case\_when(EthGrpGB = 1 & bmival >=30 ~ 'Obese', EthGrpGB = 1 & bmival < 30 ~ "Not Obese", EthGrpGB = 2 & bmival >=24 ~ 'Obese', EthGrpGB = 2 & bmival < 24 ~ "Not Obese", EthGrpGB = 3 & bmival >=28 ~ 'Obese', EthGrpGB = 3 & bmival < 28 ~ "Not Obese", EthGrpGB = 4 & bmival >=27 ~ 'Obese', EthGrpGB = 4 & bmival < 27 ~ "Not Obese"))  
  
# converting character/numeric variables into factor variables   
FinalData$bmi\_uni <- as.factor(FinalData$bmi\_uni)  
FinalData$bmi\_eth <- as.factor(FinalData$bmi\_eth)  
FinalData$EthGrpGB <- as.factor(FinalData$EthGrpGB)

### Saving the processed data

# saving the file as a csv  
save(FinalData, file = "processed data/ProcessedData.RData")