Final Project: BS 849

Irene Park, Megan Finke, Bhavya Vadavalli, Annie Goodridge

2024-03-15

## Introduction and Setup

#Loading packages and data   
library("curl") #to pull data set

## Using libcurl 8.3.0 with Schannel

library("rjags")

## Loading required package: coda

## Linked to JAGS 4.3.1

## Loaded modules: basemod,bugs

library("coda")  
library("formatR")

## Warning: package 'formatR' was built under R version 4.3.3

library("dplyr")

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

tics.data.2021 <- read.csv("tics.data.2021.csv")

Before we began our analysis, we cleaned and manipulated the data set so that it was usable for the JAGS code. We recoded Yes/No dichotomous variables such that “Yes” was 1, and “No” was 0. For MC.Asprin, we removed NA values. For other variables with NA values, we chose to handle them through imputation at later stages rather than directly removing NA values.

# Select baseline observations and variables of interest  
baseline\_data <- tics.data.2021 %>%  
 filter(!is.na(TICS01)) %>%  
 select(ID, ptype, Age.at.Enrollment, Age01, BMI, SH.Ever.Smoked., MC.Aspirin, MC.Stroke, MC.Diabetes.Mellitus, MC.HTN, MC.Coronary.Artery.Disease, MC.Cancer, MC.Heart.Attack, Years.of.Education, TICS01)  
  
# Recode "yes" and "no" variables to 1 and 0, handling missing values  
baseline\_data <- baseline\_data %>%  
 mutate(SH.Ever.Smoked. = if\_else(SH.Ever.Smoked. == "Yes", 1, if\_else(SH.Ever.Smoked. == "No", 0, NA\_integer\_)),  
 MC.Aspirin = if\_else(MC.Aspirin == "Yes", 1, if\_else(MC.Aspirin == "No", 0, NA\_integer\_)),  
 MC.Stroke = if\_else(MC.Stroke == "Yes", 1, if\_else(MC.Stroke == "No", 0, NA\_integer\_)),  
 MC.Diabetes.Mellitus = if\_else(MC.Diabetes.Mellitus == "Yes", 1, if\_else(MC.Diabetes.Mellitus == "No", 0, NA\_integer\_)),  
 MC.HTN = if\_else(MC.HTN == "Yes", 1, if\_else(MC.HTN == "No", 0, NA\_integer\_)),  
 MC.Coronary.Artery.Disease = if\_else(MC.Coronary.Artery.Disease == "Yes", 1, if\_else(MC.Coronary.Artery.Disease == "No", 0, NA\_integer\_)),  
 MC.Cancer = if\_else(MC.Cancer == "Yes", 1, if\_else(MC.Cancer == "No", 0, NA\_integer\_)),  
 MC.Heart.Attack = if\_else(MC.Heart.Attack == "Yes", 1, if\_else(MC.Heart.Attack == "No", 0, NA\_integer\_)))  
  
# Remove observations where MC.Aspirin is NA  
baseline\_data<- baseline\_data[!is.na(baseline\_data$MC.Aspirin), ]  
  
head(baseline\_data)

## ID ptype Age.at.Enrollment Age01 BMI SH.Ever.Smoked. MC.Aspirin  
## 1 1 1 64 70 22.96386 0 1  
## 2 2 1 70 76 NA 0 0  
## 3 3 1 69 74 30.58906 0 0  
## 4 4 1 64 69 21.99409 0 1  
## 5 5 1 61 67 25.55509 0 1  
## 6 6 1 77 82 20.17747 1 0  
## MC.Stroke MC.Diabetes.Mellitus MC.HTN MC.Coronary.Artery.Disease MC.Cancer  
## 1 0 0 1 0 1  
## 2 0 0 1 0 0  
## 3 0 0 1 0 1  
## 4 0 0 1 0 1  
## 5 0 0 0 0 0  
## 6 0 0 0 0 1  
## MC.Heart.Attack Years.of.Education TICS01  
## 1 0 16.0 12  
## 2 0 19.0 17  
## 3 0 17.0 15  
## 4 0 19.0 11  
## 5 0 19.5 17  
## 6 0 18.0 13

# Using summary() function  
summary(baseline\_data)

## ID ptype Age.at.Enrollment Age01   
## Min. : 1.0 Min. :0.0000 Min. :45.0 Min. :47.00   
## 1st Qu.:212.5 1st Qu.:0.0000 1st Qu.:66.0 1st Qu.:70.00   
## Median :466.0 Median :1.0000 Median :71.0 Median :76.00   
## Mean :455.3 Mean :0.6456 Mean :70.7 Mean :74.96   
## 3rd Qu.:677.5 3rd Qu.:1.0000 3rd Qu.:76.0 3rd Qu.:80.00   
## Max. :912.0 Max. :1.0000 Max. :90.0 Max. :95.00   
##   
## BMI SH.Ever.Smoked. MC.Aspirin MC.Stroke   
## Min. :17.78 Min. :0.0000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:24.01 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :26.48 Median :1.0000 Median :0.0000 Median :0.00000   
## Mean :27.25 Mean :0.5705 Mean :0.4677 Mean :0.02635   
## 3rd Qu.:29.82 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :51.18 Max. :1.0000 Max. :1.0000 Max. :1.00000   
## NA's :38   
## MC.Diabetes.Mellitus MC.HTN MC.Coronary.Artery.Disease  
## Min. :0.00000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.00000 Median :0.0000 Median :0.00000   
## Mean :0.07642 Mean :0.3531 Mean :0.05929   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.0000 Max. :1.00000   
##   
## MC.Cancer MC.Heart.Attack Years.of.Education TICS01   
## Min. :0.0000 Min. :0.00000 Min. : 8.00 Min. : 2.0   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:14.00 1st Qu.:11.0   
## Median :0.0000 Median :0.00000 Median :16.00 Median :14.0   
## Mean :0.2516 Mean :0.05007 Mean :15.79 Mean :13.9   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:18.00 3rd Qu.:16.0   
## Max. :1.0000 Max. :1.00000 Max. :26.00 Max. :26.0   
##

# Using colSums() and is.na()  
missing\_values <- colSums(is.na(baseline\_data))  
missing\_values

## ID ptype   
## 0 0   
## Age.at.Enrollment Age01   
## 0 0   
## BMI SH.Ever.Smoked.   
## 38 0   
## MC.Aspirin MC.Stroke   
## 0 0   
## MC.Diabetes.Mellitus MC.HTN   
## 0 0   
## MC.Coronary.Artery.Disease MC.Cancer   
## 0 0   
## MC.Heart.Attack Years.of.Education   
## 0 0   
## TICS01   
## 0

# Assuming 'data\_longitudinal' is your longitudinal dataset  
  
warnings()