Project 3: Evaluating the Framingham Risk Score Model in a Simulated Population Using NHANES

Mavis

```
                     Stratified by SEX
                      1                 2                 p      test
  n                      1110              1468
  CVD (mean (SD))         0.32 (0.47)       0.16 (0.37)    <0.001
  TIMECVD (mean (SD))  7226.18 (2402.62) 7952.63 (1830.88) <0.001
  SEX (mean (SD))         1.00 (0.00)       2.00 (0.00)    <0.001
  TOTCHOL (mean (SD))   226.34 (41.49)    246.22 (45.91)   <0.001
  AGE (mean (SD))        60.08 (8.23)      60.62 (8.41)     0.102
  SYSBP (mean (SD))     138.90 (21.05)    140.02 (23.74)    0.215
  DIABP (mean (SD))      81.88 (11.41)     80.33 (11.08)    0.001
  CURSMOKE (mean (SD))    0.39 (0.49)       0.31 (0.46)    <0.001
  DIABETES (mean (SD))    0.09 (0.28)       0.07 (0.25)     0.049
  BPMEDS (mean (SD))      0.11 (0.32)       0.18 (0.38)    <0.001
  HDLC (mean (SD))       43.58 (13.36)     53.03 (15.69)   <0.001
  BMI (mean (SD))        26.21 (3.49)      25.55 (4.25)    <0.001
```

```
[1] 2578   14
```

```
[1] 2539   13
```

```
                     Stratified by SEX
                      1                  2                  p      test
  n                       4557               4697
  SEQN (mean (SD))     98363.83 (2677.38) 98296.19 (2665.73)  0.223
  SYSBP (mean (SD))      122.49 (18.71)     120.20 (21.09)   <0.001
  SEX (mean (SD))          1.00 (0.00)        2.00 (0.00)    <0.001
  AGE (mean (SD))         34.12 (25.75)      34.55 (25.25)    0.419
  BMI (mean (SD))         26.16 (7.63)       26.98 (8.80)    <0.001
```

HDLC (mean (SD))	49.57 (13.53)	57.01 (14.94)	<0.001
CURSMOKE (mean (SD))	0.21 (0.41)	0.14 (0.35)	<0.001
BPMEDS (mean (SD))	0.84 (0.37)	0.86 (0.35)	0.334
TOTCHOL (mean (SD))	176.68 (40.38)	182.94 (40.59)	<0.001
DIABETES (mean (SD))	0.11 (0.31)	0.09 (0.29)	0.001

Code Appendix

```
library(riskCommunicator)
library(tidyverse)
library(tableone)
source("C:/Users/xliang34/OneDrive - Brown University/2023/missing_heatmap_function.R")
knitr::opts_chunk$set(warning = FALSE, message = FALSE,
                      echo = FALSE, fig.align = "center")
data("framingham")

# The Framingham data has been used to create models for cardiovascular risk.
# The variable selection and model below are designed to mimic the models used
# in the paper General Cardiovascular Risk Profile for Use in Primary Care
# This paper is available (cvd_risk_profile.pdf) on Canvas.

framingham_df <- framingham %>% select(c(CVD, TIMECVD, SEX, TOTCHOL, AGE,
                                         SYSBP, DIABP, CURSMOKE, DIABETES, BPMEDS,
                                         HDLC, BMI))
framingham_df <- na.omit(framingham_df)

CreateTableOne(data=framingham_df, strata = c("SEX"))

# Get blood pressure based on whether or not on BPMEDS
framingham_df$SYSBP_UT <- ifelse(framingham_df$BPMEDS == 0,
                                 framingham_df$SYSBP, 0)
framingham_df$SYSBP_T <- ifelse(framingham_df$BPMEDS == 1,
                                framingham_df$SYSBP, 0)

# Looking at risk within 15 years - remove censored data
dim(framingham_df)
framingham_df <- framingham_df %>%
  filter(!(CVD == 0 & TIMECVD <= 365*15)) %>%
  select(-c(TIMECVD))
dim(framingham_df)

# Filter to each sex
framingham_df_men <- framingham_df %>% filter(SEX == 1)
framingham_df_women <- framingham_df %>% filter(SEX == 2)

# Fit models with log transforms for all continuous variables
mod_men <- glm(CVD~log(HDLC)+log(TOTCHOL)+log(AGE)+log(SYSBP_UT+1)+
```

```
                log(SYSBP_T+1)+CURSMOKE+DIABETES,
      data= framingham_df_men, family= "binomial")


mod_women <- glm(CVD~log(HDLC)+log(TOTCHOL)+log(AGE)+log(SYSBP_UT+1)+
                   log(SYSBP_T+1)+CURSMOKE+DIABETES,
                 data= framingham_df_women, family= "binomial")


# The NHANES data here finds the same covariates among this national survey data
library(nhanesA)

# blood pressure, demographic, bmi, smoking, and hypertension info
bpx_2017 <- nhanes("BPX_J") %>%
  select(SEQN, BPXSY1 ) %>%
  rename(SYSBP = BPXSY1)
demo_2017 <- nhanes("DEMO_J") %>%
  select(SEQN, RIAGENDR, RIDAGEYR) %>%
  rename(SEX = RIAGENDR, AGE = RIDAGEYR)
bmx_2017 <- nhanes("BMX_J") %>%
  select(SEQN, BMXBMI) %>%
  rename(BMI = BMXBMI)
smq_2017 <- nhanes("SMQ_J") %>%
  mutate(CURSMOKE = case_when(SMQ040 %in% c(1,2) ~ 1,
                              SMQ040 == 3 ~ 0,
                              SMQ020 == 2 ~ 0)) %>%
  select(SEQN, CURSMOKE)
bpq_2017 <- nhanes("BPQ_J") %>%
  mutate(BPMEDS = ifelse(BPQ050A == 1, 1, 0)) %>%
  select(SEQN, BPMEDS)
tchol_2017 <- nhanes("TCHOL_J") %>%
  select(SEQN, LBXTC) %>%
  rename(TOTCHOL = LBXTC)
hdl_2017 <- nhanes("HDL_J") %>%
  select(SEQN, LBDHDD) %>%
  rename(HDLC = LBDHDD)
diq_2017 <- nhanes("DIQ_J") %>%
  mutate(DIABETES = case_when(DIQ010 == 1 ~ 1,
                              DIQ010 %in% c(2,3) ~ 0,
                              TRUE ~ NA)) %>%
  select(SEQN, DIABETES)
```

```r
# Join data from different tables
df_2017 <- bpx_2017 %>%
  full_join(demo_2017, by = "SEQN") %>%
  full_join(bmx_2017, by = "SEQN") %>%
  full_join(hdl_2017, by = "SEQN") %>%
  full_join(smq_2017, by = "SEQN") %>%
  full_join(bpq_2017, by = "SEQN") %>%
  full_join(tchol_2017, by = "SEQN") %>%
  full_join(diq_2017, by = "SEQN")

CreateTableOne(data = df_2017, strata = c("SEX"))
```