Denis Ostroushko - HW3

Introduction

Imputation and Variable Processing

Drop Hisp

BMI, BL.Cig.Day, , BL.Drks.Day, , N.living.kids impute with medians like N.prev.preg, Birthweight

Use.Alc, Drug.Add impute with modes like Race_ethnicity, Use.Tob

Problem 1

fit all models below

```
logistic_regression_a <-
  glm(
    I(data$Group == "T") %>% as.numeric() ~

        Race_ethnicity + Public.Asstce +
        Use.Tob + Live.PTB +

        N.prev.preg + BL.GE + BL..BOP + BL..PD.4 + BL..CAL.3,

    data = data,
    family = "binomial"
  )

logistic_regression_b <-
  glm(
    I(data$Group == "T") %>% as.numeric() ~
```

```
        Race_ethnicity + Public.Asstce +
        Use.Tob + Live.PTB +

        poly(N.prev.preg,2) +
        poly(BL.GE,5) +
        poly(BL..BOP,5) +
        poly(BL..PD.4,5) +
        poly(BL..CAL.3, 2),

    data = data,
    family = "binomial"
  )

logistic_regression_c_full <-
  glm(
    I(data$Group == "T") %>% as.numeric() ~

        (Race_ethnicity + Public.Asstce +
        Use.Tob + Live.PTB +

        N.prev.preg + BL.GE + BL..BOP + BL..PD.4 + BL..CAL.3)^2,

    data = data,
    family = "binomial"
  )


logistic_regression_c_lower <- glm(I(data$Group == "T") %>% as.numeric() ~ 1, data, family

logistic_regression_c <- MASS::stepAIC(logistic_regression_c_lower, direction = "forward",
                                       scope = list(upper = logistic_regression_c_full,
                                                    lower = logistic_regression_c_lower))

data2 = data %>% select(- PID, - Birth.outcome, - GA.at.outcome, -Preg.ended...37.wk, -Bir
  mutate(Race_ethnicity = as.factor(Race_ethnicity))

logistic_regression_d_full <-
  glm(
    I(data$Group == "T") %>% as.numeric() ~ .,
    data = data2,
    family = "binomial"
```

```
  )


logistic_regression_c_lower <- glm(I(data$Group == "T") %>% as.numeric() ~ 1, data, family

logistic_regression_d <- MASS::stepAIC(logistic_regression_c_lower, direction = "forward",
                                       scope = list(upper = logistic_regression_d_full,
                                                    lower = logistic_regression_c_lower))

rf = randomForest(
  Group ~ .,
  data = data %>% select(-Drug.Add),
  ntree = 1000
)
```