Obesity

Darwin Stahlback , Ani Lamichhane, Karen Vuong

12/15/2021

```
library(tidyverse)
library(mosaic)
library(mosaicData)
library(lmtest)
library(skimr)
library(broom)
library(readr)
library(dplyr)
library(janitor)
library(dplyr)
```

```
#import data
obesity <- read_csv("Group-M-Obesity-Data.csv")
```

Data wrangling:

```
obesity <- obesity %>%  clean_names() #all column names to lower case
obesity$caec <- tolower(obesity$caec) #all column values to lower case
obesity$mtrans <- tolower(obesity$mtrans)#all column values to lower case
obesity$n_obeyesdad  <- tolower(obesity$n_obeyesdad)#all column values to lower case



obesity <- obesity %>%
  mutate(massbodyindex = obesity$weight/(obesity$height*obesity$height)) #this takes original n_obeyesd
obesity$is_obese = ifelse(obesity$massbodyindex<30,0,1) # creates binary if obese 0 = no / 1 = yes
is.num <- sapply(obesity, is.numeric)
obesity[is.num] <- lapply(obesity[is.num], round, 3) # format all numeric column to three places
```

```
#renaming variables
colnames(obesity)[colnames(obesity) == "family_history_with_overweight"] = "family_history"

colnames(obesity)[colnames(obesity) == "favc"] = "food_between_meals"

colnames(obesity)[colnames(obesity) == "caec"] = "main_meals"

colnames(obesity)[colnames(obesity) == "scc"] = "monitor_calories"

colnames(obesity)[colnames(obesity) == "calc"] = "alcohol"
```

```r
colnames(obesity)[colnames(obesity) == "mtrans"] = "transportation"

names(obesity)

obesityuse = obesity%>%
  select(gender, age, height, weight, food_between_meals, main_meals, smoke, monitor_calories, alcohol,
```

Variable Selection Multiple Logistic Regression:

```r
#does not have weight because it won't work with it
logmvarselcm1 = glm(is_obese ~ gender+age+height+food_between_meals+main_meals+smoke+monitor_calories+al
summary(logmvarselcm1)
```

```r
logmvarselcm2 = glm(is_obese ~ gender+age+height+food_between_meals+main_meals+smoke+monitor_calories+tr
summary(logmvarselcm2)
```

```r
logmvarselcm3 = glm(is_obese ~ gender+age+height+food_between_meals+main_meals+monitor_calories+transpo
summary(logmvarselcm3)
```

Variable Selection Simple Logistic Regression:

```r
sm1 = glm(is_obese ~ gender, data = obesityuse, family = binomial)

sm2 = glm(is_obese ~ age, data = obesityuse, family = binomial)

sm3 = glm(is_obese ~ height, data = obesityuse, family = binomial)

sm4 = glm(is_obese ~ food_between_meals, data = obesityuse, family = binomial)

sm5 = glm(is_obese ~ main_meals, data = obesityuse, family = binomial)

sm6 = glm(is_obese ~ smoke, data = obesityuse, family = binomial)

sm7 = glm(is_obese ~ monitor_calories, data = obesityuse, family = binomial)

sm8 = glm(is_obese ~ alcohol, data = obesityuse, family = binomial)

sm9 = glm(is_obese ~ transportation, data = obesityuse, family = binomial)

AIC(sm1, sm2, sm3, sm4, sm5, sm6, sm7, sm8, sm9)
```

Model:

```r
m5 = glm(is_obese ~ main_meals, data = obesityuse, family = binomial)
summary(m5)
exp(coef(m5))
```

Odds Space: yhat_always = 0.1777815 yhat_frequently = 0.1777815 + 0.1922997 = 0.3700812 yhat_no = 0.1777815 + 0.2295809 = 0.4073624 yhat_sometimes = 0.1777815 + 6.647228 = 6.82501

Condition:

Independence: Does not violate independence because the rows do not include time or spatial units.

Randomness: Obesity is a random factor. You don't really know if someone will be obese or not.

Linearity of the Logit:

We don't know how to check for linearity of the logit when our variable is categorical. Also we were trying to run multiple logistic regression and we also don't know how to write a model or check for linearity of the logit for multiple logistic regression.

Coefficient interpretation:

People that always have a main meal are associated with multiplying the odds of having obesity by a factor of 0.1777778. (the odds of having obesity goes down.)

People that frequently have a main meal are associated with multiplying the odds of having obesity by a factor of 0.1923080 (the odds of having obesity goes down.)

People that don't have a main meal are associated with multiplying the odds of having obesity by a factor of 0.2295919 (the odds of having obesity goes down.)

People that sometimes have a main meal are associated with multiplying the odds of having obesity by a factor of 6.6470952 (the odds of having obesity goes up.)