STA-363-Lab-5

Phoebe Yan & Rickey huang

3/10/2021

```r
#import the data
source("http://www.openintro.org/stat/data/cdc.R")
```

```r
#change the variable
cdc$wtdesire<- cdc$weight -cdc$wtdesire
```

Question 1

Based on this, how many models do you think we fit in Stage 1 (Step 1) of BSS? Hint: Think carefully here.
Some of the variables are categorical with more than two levels. Each level counts as a possible X.

We think when we fit the models in the Stage 1 of BSS, we would fit 11 models with 15 possible Xs. Each
numeric variables can be served as a possible X to fit a model. There are 3 numeric variables among all
variables. For the categorical variables with only 2 levels, we can use one level between the two to represent
the variable and set it to 1 when the indivial is in this level and 0 otherwise. There are 4 categorical variables
in this dataset. In addition, the variable *genhlth* is a categorical variable with 5 levels, which are *poor*, *fair*,
excellent, *good*, and *very good*, and each of them can served as a separate variable choice for the model. Hence,
we have $3 + 4 + 5 = 11$ models in total we can fit during the first step of BSS.