Causal Inference Final Project: Effect of Smoking on 10-year Development of Cardiovascular Disease

Daniel Saunders

November 25, 2018

Data Exploration

```
fhs = read.csv('framingham.csv', header = T)
head(fhs)
```

```
##   male age education currentSmoker cigsPerDay BPMeds prevalentStroke
## 1    1  39         4             0          0      0               0
## 2    0  46         2             0          0      0               0
## 3    1  48         1             1         20      0               0
## 4    0  61         3             1         30      0               0
## 5    0  46         3             1         23      0               0
## 6    0  43         2             0          0      0               0
##   prevalentHyp diabetes totChol sysBP diaBP   BMI heartRate glucose
## 1            0        0     195 106.0    70 26.97        80      77
## 2            0        0     250 121.0    81 28.73        95      76
## 3            0        0     245 127.5    80 25.34        75      70
## 4            1        0     225 150.0    95 28.58        65     103
## 5            0        0     285 130.0    84 23.10        85      85
## 6            1        0     228 180.0   110 30.30        77      99
##   TenYearCHD
## 1          0
## 2          0
## 3          0
## 4          1
## 5          0
## 6          0
```

```
summary(fhs)
```

```
##       male             age          education     currentSmoker
##  Min.   :0.0000   Min.   :32.00   Min.   :1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:42.00   1st Qu.:1.000   1st Qu.:0.0000
##  Median :0.0000   Median :49.00   Median :2.000   Median :0.0000
##  Mean   :0.4292   Mean   :49.58   Mean   :1.979   Mean   :0.4941
##  3rd Qu.:1.0000   3rd Qu.:56.00   3rd Qu.:3.000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :70.00   Max.   :4.000   Max.   :1.0000
##                                   NA's   :105
##    cigsPerDay         BPMeds        prevalentStroke     prevalentHyp
##  Min.   : 0.000   Min.   :0.00000   Min.   :0.000000   Min.   :0.0000
##  1st Qu.: 0.000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.0000
##  Median : 0.000   Median :0.00000   Median :0.000000   Median :0.0000
##  Mean   : 9.006   Mean   :0.02962   Mean   :0.005896   Mean   :0.3106
##  3rd Qu.:20.000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:1.0000
##  Max.   :70.000   Max.   :1.00000   Max.   :1.000000   Max.   :1.0000
##  NA's   :29       NA's   :53
##     diabetes          totChol          sysBP           diaBP      
##  Min.   :0.00000   Min.   :107.0   Min.   : 83.5   Min.   : 48.0  
##  1st Qu.:0.00000   1st Qu.:206.0   1st Qu.:117.0   1st Qu.: 75.0  
##  Median :0.00000   Median :234.0   Median :128.0   Median : 82.0  
##  Mean   :0.02571   Mean   :236.7   Mean   :132.4   Mean   : 82.9  
##  3rd Qu.:0.00000   3rd Qu.:263.0   3rd Qu.:144.0   3rd Qu.: 90.0  
##  Max.   :1.00000   Max.   :696.0   Max.   :295.0   Max.   :142.5  
##                    NA's   :50                                     
##       BMI          heartRate         glucose         TenYearCHD    
##  Min.   :15.54   Min.   : 44.00   Min.   : 40.00   Min.   :0.0000  
##  1st Qu.:23.07   1st Qu.: 68.00   1st Qu.: 71.00   1st Qu.:0.0000  
##  Median :25.40   Median : 75.00   Median : 78.00   Median :0.0000  
##  Mean   :25.80   Mean   : 75.88   Mean   : 81.96   Mean   :0.1519  
##  3rd Qu.:28.04   3rd Qu.: 83.00   3rd Qu.: 87.00   3rd Qu.:0.0000  
##  Max.   :56.80   Max.   :143.00   Max.   :394.00   Max.   :1.0000  
##  NA's   :19      NA's   :1        NA's   :388                      
```

```r
table(fhs$TenYearCHD)
```

```
## 
##    0    1 
## 3596  644 
```

```r
table(fhs$currentSmoker)
```

```
## 
##    0    1 
## 2145 2095 
```

```r
table(fhs$cigsPerDay)
```

```
## 
##    0    1    2    3    4    5    6    7    8    9   10   11   12   13   14 
## 2145   67   18  100    9  121   18   12   11  130  143    5    3    3    2 
##   15   16   17   18   19   20   23   25   29   30   35   38   40   43   45 
##  210    3    7    8    2  734    6   55    1  218   22    1   80   56    3 
##   50   60   70 
##    6   11    1 
```

```r
dim(fhs)
```

```
## [1] 4240   16
```

Causal Roadmap

Step 0: Specify the Scientific Question

What is the effect of smoking on the ten-year development of Cardiovascular Disease? The target population is white middle-class men and women aged 30 to 62 (at baseline) in Framingham, Massachusetts. **Note: Professor Balzer's comment on our project description Google Document suggests this might be inadequate. Can we claim that our results generalize outside of Framingham? Why?**

Step 1: Specify a Causal Model