R Notebook

```r
# load data set
heartattack <- read.csv("heart_attack_prediction_dataset.csv", header=T)

head(heartattack)
```

```
##   Patient.ID Age    Sex Cholesterol Blood.Pressure Heart.Rate Diabetes
## 1    BMW7812  67   Male         208         158/88         72        0
## 2    CZE1114  21   Male         389         165/93         98        1
## 3    BNI9906  21 Female         324         174/99         72        1
## 4    JLN3497  84   Male         383        163/100         73        1
## 5    GFO8847  66   Male         318          91/88         93        1
## 6    ZOO7941  54 Female         297         172/86         48        1
##   Family.History Smoking Obesity Alcohol.Consumption Exercise.Hours.Per.Week
## 1              0       1       0                   0                4.168189
## 2              1       1       1                   1                1.813242
## 3              0       0       0                   0                2.078353
## 4              1       1       0                   1                9.828130
## 5              1       1       1                   0                5.804299
## 6              1       1       0                   1                0.625008
##       Diet Previous.Heart.Problems Medication.Use Stress.Level
## 1  Average                       0              0            9
## 2 Unhealthy                      1              0            1
## 3   Healthy                      1              1            9
## 4   Average                      1              0            9
## 5 Unhealthy                      1              0            6
## 6 Unhealthy                      1              1            2
##   Sedentary.Hours.Per.Day Income      BMI Triglycerides
## 1                6.615001 261404 31.25123           286
## 2                4.963459 285768 27.19497           235
## 3                9.463426 235282 28.17657           587
## 4                7.648981 125640 36.46470           378
## 5                1.514821 160555 21.80914           231
## 6                7.798752 241339 20.14684           795
##   Physical.Activity.Days.Per.Week Sleep.Hours.Per.Day   Country      Continent
## 1                               0                   6 Argentina South America
## 2                               1                   7    Canada North America
## 3                               4                   4    France        Europe
## 4                               3                   4    Canada North America
## 5                               1                   5  Thailand          Asia
## 6                               5                  10   Germany        Europe
##              Hemisphere Heart.Attack.Risk
## 1 Southern Hemisphere                   0
## 2 Northern Hemisphere                   0
## 3 Northern Hemisphere                   0
## 4 Northern Hemisphere                   0
## 5 Northern Hemisphere                   0
```

```
## 6 Northern Hemisphere                    1
```

Find recommended sample size for this study

```r
# calculate min sample size needed
pop_size <- nrow(heartattack) # 8763

# using 95% CI, find n for worst case scenario: p = 0.5
MOE <- 0.05
z <- 1.96
p_guess <- 0.5

# if N is large enough to ignore FPC
n_0 = ceiling( ((2*z)^2*(0.5)*(0.5)) / (MOE^2)) # 1537

# since we know N = 8763, using FPC
n = ceiling( n_0 / (1 + (n_0/pop_size)) ) # 1308
```

Selecting Sample

```r
# set seed
set.seed(2023)

# take SRS of n = 1308
SRS.index <- sample.int(pop_size, n, replace=F)
heartsample <- heartattack[SRS.index, ]
head(heartsample)
```

```
##         Patient.ID Age    Sex Cholesterol Blood.Pressure Heart.Rate Diabetes
## 5584      KCR6223  87   Male         123         119/91         57        1
## 1992      BQZ1708  87 Female         256        107/106         70        1
## 7923      HIX2091  51   Male         230         161/95        102        1
## 7724      MCG9175  41   Male         309         172/60        101        1
## 866       FPU3089  20   Male         376         123/97         55        1
## 7361      LVB4887  76   Male         270         168/91        108        1
##         Family.History Smoking Obesity Alcohol.Consumption Exercise.Hours.Per.Week
## 5584                 1       1       0                   1               18.146292
## 1992                 1       1       0                   0                5.077104
## 7923                 0       1       1                   0                3.814575
## 7724                 0       1       1                   1                7.530109
## 866                  0       1       0                   1               10.829282
## 7361                 1       1       1                   0               15.745186
##             Diet Previous.Heart.Problems Medication.Use Stress.Level
## 5584   Unhealthy                       1              1            6
## 1992   Unhealthy                       1              0            7
## 7923     Average                       1              0            1
## 7724     Healthy                       0              0            9
## 866    Unhealthy                       1              0           10
## 7361   Unhealthy                       0              0            2
##         Sedentary.Hours.Per.Day Income      BMI Triglycerides
## 5584              11.913423247 224496 29.99098           556
## 1992               4.736410347 174159 32.03624           568
## 7923               3.180108757 233545 33.96761           397
```

```
## 7724                  9.291229691 226143 32.99921           434
## 866                   0.001529091 293544 34.06575           210
## 7361                  4.347936688 267371 39.02180           604
##      Physical.Activity.Days.Per.Week Sleep.Hours.Per.Day        Country
## 5584                               3                   7          Italy
## 1992                               7                   8        Germany
## 7923                               2                  10  United States
## 7724                               3                   6   South Africa
## 866                                5                   9         Canada
## 7361                               1                   4         France
##          Continent          Hemisphere Heart.Attack.Risk
## 5584        Europe Southern Hemisphere                 1
## 1992        Europe Northern Hemisphere                 0
## 7923 North America Northern Hemisphere                 0
## 7724        Africa Southern Hemisphere                 0
## 866  North America Northern Hemisphere                 1
## 7361        Europe Northern Hemisphere                 0
```