Homework 1

Aji John

Q1. For the scientific question of interest, what are the two primary variables in this study? What is the response variable and what is the predictor variable? What types of variables are they: quantitative or qualitative? Discrete or continuous? Nominal, ordinal, interval, or ratios? Censored (right, left, or both) or uncensored? Explain briefly.

Q1A. The two primary variables are 'atropy', and 'age'. Response variable is 'atropy', and predictor variable is 'age'. Both the variables are quantitative, 'Age' and 'Atropy' are discrete.

```
## 'data.frame':    735 obs. of  30 variables:
##  $ ptid    : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ mridate : int  120791 90192 82092 73192 111691 82292 71892 82692 80692 122191 ...
##  $ age     : int  72 81 90 72 70 72 75 75 67 70 ...
##  $ male    : int  1 0 1 1 0 1 1 1 0 0 ...
##  $ race    : int  2 2 2 1 1 4 1 2 1 1 ...
##  $ weight  : num  173 139 145 190 153 ...
##  $ height  : num  169 170 170 181 158 ...
##  $ packyrs : num  54 0 0 33 0 ...
##  $ yrsquit : int  0 0 0 17 0 21 12 0 0 0 ...
##  $ alcoh   : num  0 0.25 1.25 9.5 0.25 21 0 1 0 8 ...
##  $ physact : num  9.84 0.78 1.64 3.52 0.75 ...
##  $ chf     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ chd     : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ stroke  : int  2 0 0 0 0 0 2 0 0 0 ...
##  $ diabetes: int  0 0 0 0 0 0 1 0 0 0 ...
##  $ genhlth : int  3 2 3 2 2 2 2 3 3 3 ...
##  $ ldl     : int  135 84 115 61 148 163 101 116 124 110 ...
##  $ alb     : num  3.7 3.8 4.2 4.3 4.1 3.9 3.7 3.9 NA 3.2 ...
##  $ crt     : num  1.4 1.3 1.2 1.1 0.6 1 1 1 NA 0.6 ...
##  $ plt     : int  275 142 192 133 266 539 167 198 NA 108 ...
##  $ sbp     : num  139 146 134 147 117 146 140 171 121 89 ...
##  $ aai     : num  1.03 1.11 1.014 0.98 0.949 ...
##  $ fev     : num  1.28 2.55 2.38 2.7 2.03 ...
##  $ dsst    : int  25 51 27 43 48 40 44 35 46 47 ...
##  $ atrophy : int  20 43 35 32 27 18 38 20 40 36 ...
##  $ whgrd   : int  2 2 1 2 1 1 2 1 2 1 ...
##  $ numinf  : int  1 3 2 1 0 2 0 0 1 0 ...
##  $ volinf  : num  7.4613 0.1414 0.1885 0.0419 0 ...
##  $ obstime : int  2110 1841 1853 1873 2131 1851 1886 1847 1867 2096 ...
##  $ death   : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## # A tibble: 735 x 2
```

```
##     ptid count
##    <int> <int>
##  1     1     1
##  2     2     1
##  3     3     1
##  4     4     1
##  5     5     1
##  6     6     1
##  7     7     1
##  8     8     1
##  9     9     1
## 10    10     1
## # ... with 725 more rows
```

2.What is the population of interest for this study? What is the sample? What is the size of the sample? Are there any individuals in the sample who have missing data on smoking history? If so, provide the participant identification numbers for any study individuals who have missing data on smoking history.