# zjin26_mini-assign4

Vincent Jin

2023-04-03

==========

## Mini-assignment # 4

==========

## (1) calculate and store the min, max, mean, and standard deviation of all numerical columns for the following data.table:

```
library(data.table)
set.seed(100)
a = sample(65:110,1000,replace=T)
b = sample(c('M','F','F'),1000,replace=T)
c = sample(c('W','W','W','B','B','A','O',NA),1000,replace=T)
d = sample(100:5000,1000)*a
e = runif(1000,0,1)*(a/110)
dt = data.table(age = a, sex = b, race = c, cost = d, risk = e)
```

```
col <- names(dt)

for (i in col) {
  if (typeof(dt[[i]]) != "character") {
    cat(paste("the min, max, mean, sd of", i, "is :\n"))
    print(min(dt[[i]]))
    print(max(dt[[i]]))
    print(mean(dt[[i]]))
    print(sd(dt[[i]]))
  }
}
```

```
## the min, max, mean, sd of age is :
## [1] 65
## [1] 110
## [1] 87.577
## [1] 13.02896
## the min, max, mean, sd of cost is :
```

```
## [1] 7560
## [1] 535408
## [1] 217375.1
## [1] 131361.2
## the min, max, mean, sd of risk is :
## [1] 0.0006583922
## [1] 0.9985355
## [1] 0.3904799
## [1] 0.2360103
```

## (2) print mean and standard deviation (2 decimals) of the numerical columns as below (replace XXX with actual variables):

## The study population has an average age of XXX years with a standard deviation of XXX years

## The study population has an average cost of XXX dollars with a maximum of XXX dollars

## The total risk ranged between minimum of XXX and maximum of XXX

```
print(paste("The study population has an average age of", round(mean(dt$age), 2), "years with a standar
```

```
## [1] "The study population has an average age of 87.58 years with a standard deviation of 13.03 years
```

```
print(paste("The study population has an average cost of", round(mean(dt$cost), 2), "dollars with a sta
```

```
## [1] "The study population has an average cost of 217375.11 dollars with a standard deviation of 1313
```

```
print(paste("The total risk ranged between minimum of", round(min(dt$risk), 2), "and maximum of", round
```

```
## [1] "The total risk ranged between minimum of 0 and maximum of 1"
```

## (3) calculate the mean age, cost, and risk of each race as well as populations with missing race separately

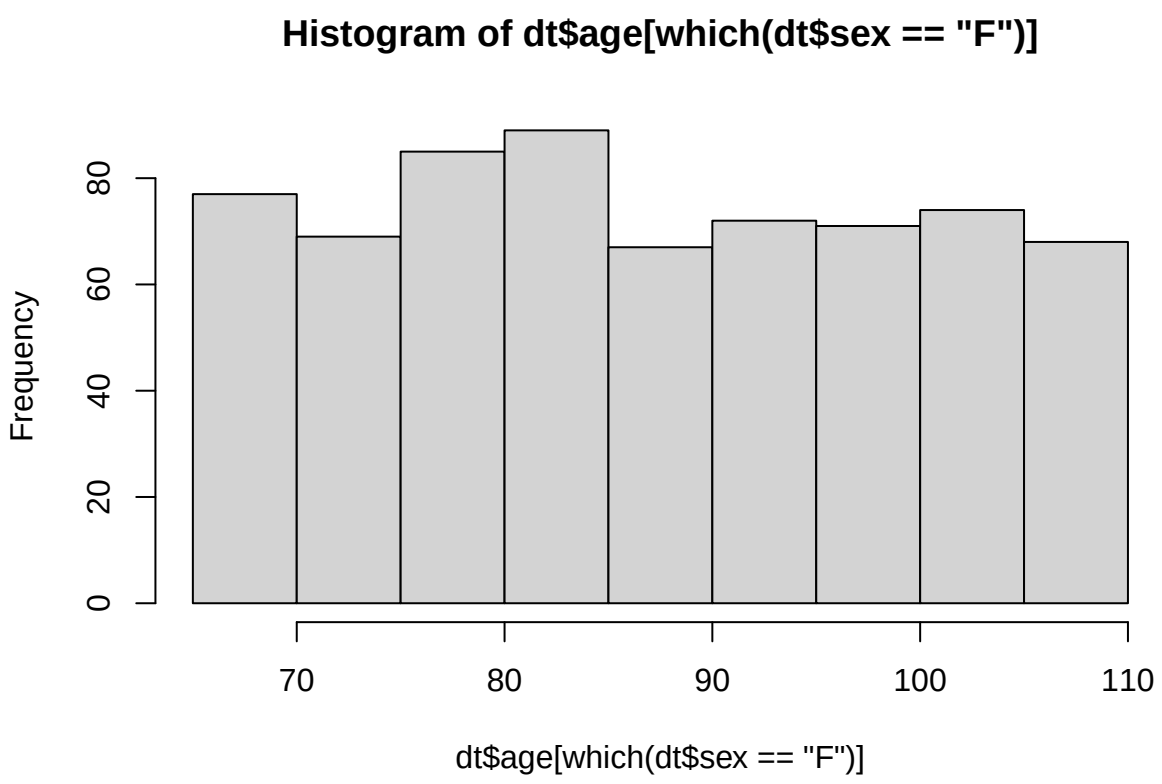
## try using the short/alternate built-in data.table syntax (although base R syntax is also acceptable)

```
col <- c("age", "cost", "risk")
race <- unique(dt$race)
for (i in col) {
  for (r in race) {
    if (is.na(r) != TRUE) {
      print(paste("the mean", i, "of race", r, "is: "))
      print(mean(dt[[i]][which(dt$race == r)]))
    } else {
      print(paste("the mean", i, "of race", r, "is: "))
      print(mean(dt[[i]][which(is.na(dt$race))]))
    }
  }
}
```

```
## [1] "the mean age of race W is: "
## [1] 88.24936
## [1] "the mean age of race O is: "
## [1] 85.59375
## [1] "the mean age of race B is: "
## [1] 88.94737
## [1] "the mean age of race A is: "
## [1] 86.16364
## [1] "the mean age of race NA is: "
## [1] 85.9918
## [1] "the mean cost of race W is: "
## [1] 224195.4
## [1] "the mean cost of race O is: "
## [1] 197264.9
## [1] "the mean cost of race B is: "
## [1] 224066.5
## [1] "the mean cost of race A is: "
## [1] 206444.7
## [1] "the mean cost of race NA is: "
## [1] 212812
## [1] "the mean risk of race W is: "
## [1] 0.3751833
## [1] "the mean risk of race O is: "
## [1] 0.4171963
## [1] "the mean risk of race B is: "
## [1] 0.4138082
## [1] "the mean risk of race A is: "
## [1] 0.3724946
## [1] "the mean risk of race NA is: "
## [1] 0.3807112
```

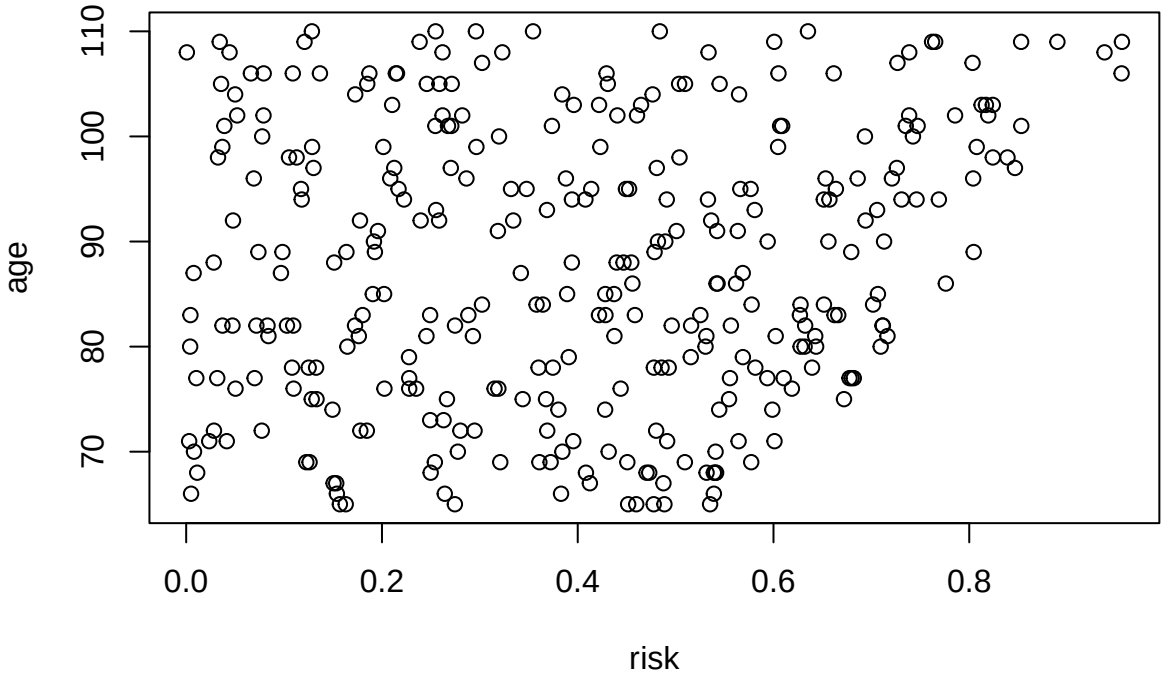
## (4) show the histogram of age for the female population

```
hist(dt$age[which(dt$sex == "F")])
```




## (5) show the scatter plot of age vs. risk for the male population

```
plot(age ~ risk, data = dt[which(dt$sex == "M")])
```


age
70
80
90
100
110
0.0
0.2
0.4
0.6
0.8
risk