> # Task -1

>

> d=read.csv(file.choose())

> View(d)

> str(d) # To know the structure of the variables

'data.frame': 891 obs. of 12 variables:

$ PassengerId: int 1 2 3 4 5 6 7 8 9 10 ...

$ Survived : int 0 1 1 1 0 0 0 0 1 1 ...

$ Pclass : int 3 1 3 1 3 3 1 3 3 2 ...

$ Name : chr "Braund, Mr. Owen Harris" "Cumings, Mrs. John Bradley (Florence Briggs Thayer)" "Heikkinen, Miss. Laina" "Futrelle, Mrs. Jacques Heath (Lily May Peel)" ...

$ Sex : chr "male" "female" "female" "female" ...

$ Age : num 22 38 26 35 35 NA 54 2 27 14 ...

$ SibSp : int 1 1 0 1 0 0 0 3 0 1 ...

$ Parch : int 0 0 0 0 0 0 0 1 2 0 ...

$ Ticket : chr "A/5 21171" "PC 17599" "STON/O2. 3101282" "113803" ...

$ Fare : num 7.25 71.28 7.92 53.1 8.05 ...

$ Cabin : chr "" "C85" "" "C123" ...

$ Embarked : chr "S" "C" "S" "S" ...

> length(d)

[1] 12

> # Removing missing values

> d1=na.omit(d) # cleaned data

> View(d1)

> summary(d1)

PassengerId Survived Pclass Name

Min. : 1.0 Min. :0.0000 Min. :1.000 Length:714

1st Qu.:222.2 1st Qu.:0.0000 1st Qu.:1.000 Class :character

Median :445.0 Median :0.0000 Median :2.000 Mode :character

Mean :448.6 Mean :0.4062 Mean :2.237

3rd Qu.:677.8 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :891.0 Max. :1.0000 Max. :3.000

Sex Age SibSp Parch

Length:714 Min. : 0.42 Min. :0.0000 Min. :0.0000

Class :character 1st Qu.:20.12 1st Qu.:0.0000 1st Qu.:0.0000

Mode :character Median :28.00 Median :0.0000 Median :0.0000

Mean :29.70 Mean :0.5126 Mean :0.4314

3rd Qu.:38.00 3rd Qu.:1.0000 3rd Qu.:1.0000

Max. :80.00 Max. :5.0000 Max. :6.0000

Ticket Fare Cabin

Length:714 Min. : 0.00 Length:714

Class :character 1st Qu.: 8.05 Class :character

Mode :character Median : 15.74 Mode :character

Mean : 34.69

3rd Qu.: 33.38

Max. :512.33

Embarked

Length:714

Class :character

Mode :character

> #Task -2

> View(d1) #cleaned data

> # Summary statistics

> summary(d1)

PassengerId Survived Pclass Name

Min. : 1.0 Min. :0.0000 Min. :1.000 Length:714

1st Qu.:222.2 1st Qu.:0.0000 1st Qu.:1.000 Class :character

Median :445.0 Median :0.0000 Median :2.000 Mode :character

Mean :448.6 Mean :0.4062 Mean :2.237

3rd Qu.:677.8 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :891.0 Max. :1.0000 Max. :3.000

Sex Age SibSp Parch

Length:714 Min. : 0.42 Min. :0.0000 Min. :0.0000

Class :character 1st Qu.:20.12 1st Qu.:0.0000 1st Qu.:0.0000

Mode :character Median :28.00 Median :0.0000 Median :0.0000

Mean :29.70 Mean :0.5126 Mean :0.4314

3rd Qu.:38.00 3rd Qu.:1.0000 3rd Qu.:1.0000

Max. :80.00 Max. :5.0000 Max. :6.0000

Ticket Fare Cabin

Length:714 Min. : 0.00 Length:714

Class :character 1st Qu.: 8.05 Class :character

Mode :character Median : 15.74 Mode :character

Mean : 34.69

3rd Qu.: 33.38

Max. :512.33

Embarked

Length:714

Class :character

Mode :character

> # Standard Deviations

> sd(d1[,2]) #sd of Survived

[1] 0.4914599

> sd(d1[,3]) #sd of Pclass

[1] 0.8382499

> sd(d1[,6]) #sd of Age

[1] 14.5265

> sd(d1[,7]) #sd of SibSp

[1] 0.9297835

> sd(d1[,8]) #sd of Parch

[1] 0.8532894

> sd(d1[,10]) #sd of Fare

[1] 52.91893

>