

## R PROGRAMMING

Ragu.v  
191921031

REG NO:

1. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. What is the median?

Solution

```
> age <- c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)
```

```
> median age <- median (age)
```

```
> print (median age)
```

Output

```
[1] 25
```

2. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?

Solution

```
> age <- c (13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)
```

```
> g1 <- quantile (age, 0.25)
```

```
> g3 <- quantile (age, 0.75)
```

```
> print (g1)
```

```
25%
```

```
20.5
```

```
> print (g3)
```

```
75%
```

```
35
```

3. Load iris Dataset which is inbuilt in R .explore the dataset in terms of dimension and summary statistics

Solution

```
> library (datasets)
```

```
> data (iris)
```

```
> head (iris)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
--------------	-------------	--------------	-------------	---------

1				
---	--	--	--	--

5.1				
-----	--	--	--	--

3.5				
-----	--	--	--	--

1.4				
-----	--	--	--	--

2				
---	--	--	--	--

4.9				
-----	--	--	--	--

3.0				
-----	--	--	--	--

1.4				
-----	--	--	--	--

3				
---	--	--	--	--

4.7				
-----	--	--	--	--

3.2				
-----	--	--	--	--

1.3				
-----	--	--	--	--

4				
---	--	--	--	--

4.6				
-----	--	--	--	--

3.1				
-----	--	--	--	--

1.5				
-----	--	--	--	--

5				
---	--	--	--	--

5.0				
-----	--	--	--	--

3.6				
-----	--	--	--	--

1.4				
-----	--	--	--	--

6				
---	--	--	--	--

5.4				
-----	--	--	--	--

3.9				
-----	--	--	--	--

1.7				
-----	--	--	--	--

0.2				
-----	--	--	--	--

setosa				
--------	--	--	--	--

0.2				
-----	--	--	--	--

setosa				
--------	--	--	--	--

0.2				
-----	--	--	--	--

setosa				
--------	--	--	--	--

0.2 setosa				
------------	--	--	--	--

0.2				
-----	--	--	--	--

setosa				
--------	--	--	--	--

0.4 setosa				
------------	--	--	--	--

```
> summary (iris)
```

Sepal.Length
--------------

Min.
------

:4.300
--------

1st Qu.:5.100
---------------

Median :5.800
---------------

Mean
------

:5.843
--------

3rd Qu.: 6.400
----------------

```

Max.
:7.900
Sepal.Width
Min. :2.000
1st Qu.:2.800
Median :3.000
Mean
:3.057
3rd Qu.: 3.300
Max.
4.400
Petal.Length
Petal.Width
Min.
:1.000
Min. :0.100
1st Qu.: 1.600
1st Qu.: 0.300
Median :4.350
Median :1.300
Mean
:3.758
Mean
:1.199
3rd Qu.: 5.100
3rd Qu.: 1.800
Max.
:6.900
Max.
:2.500
Species
setosa
:50
versicolor: 50
virginica :50

```

4. Find the categorical column data and convert that to factor form, also find the number of rows for each factors in dataset.

Solution

```
> df <- data.frame(
```

```
+name=c("John", "Jane", "Bob", "Sara", "Mike", "Emma", "Tom"),
```

```
gender = c("Male", "Female", "Male", "Female", "Male", "Female", "Male"), + marital status =
c("Married", "Single", "Married", "Single", "Divorced", "Single", "Widowed")
```

```
age = c(25, 30, 35, 20, 45, 28, 33), city= c("Kolkata", "Delhi", "Chennai", "Hubali", "Mysore",
"Bombay", "Surat"),
```

```
> dfsgender <- factor (dfsgender)
```

```
> df marital status <- factor (dfSmarital status) > table (dfsgender)
```

Female Male

3

4

```
> table (df@marital status)
```

Divorced Married

1

2

Single

3

Widowed

1

5. Find mean of numeric

6. Find mean of numeric data in dataset based on Species group. and plot Bar chart (use ggplot ) to interpret same

Species	Sepal Length	Sepal Width	Petal Length	Petal Width
Setosa	5.006	3.428	1.462	0.246
Versicolour	5.936	2.770	4.250	1.326
Virginica	6.588	2.974	5.552	2.026

I. Draw a suitable plot which summaries statistical parameter of Sepal.Width based on Species group

```
> library(dplyr)
```

```
> library(ggplot2)
```

```

> df <- data.frame( Species c("Setosa", "Versicolour", "Virginia"),

Sepal.Length (5.006, 5.936, 6.500),

Sepal.Width (3.428, 2.77, 2.974), Petal.Length c(1.462, 4.25, 5.552),

Petal.Width c(0.246, 1.326, 2.026)

>df mean of 44

group by (Species)

summarise all (mean)

> ggplot (df_mean, aes(x Species, y Sepal.Width))+

gece bar (stat "identity")

labs (title="Mean Sepal Width based on Species group", x "Species", y "Mean Sepal
Width")

> ggplot (df, aes(x Species, y Sepal.Width)) +

geom boxplot() +

labs (title "Statistical parameter of Sepal Width based on Species group", x "Species",
y "Sepel Width")

> ggplot (df, aes (x Sepal.Width, fill Species))+

geom density (alpha 0.5)+

labs (title "Density plot of Sepal Width", "Sepal Width")

```

Warning messages:

1: Groups with fewer than two data points have been dropped.

2: Groups with fewer than two data points have been dropped.

3: Groups with fewer than two data points have been dropped.

4: In max(ids, na.rm TRUE):

no non-missing arguments to max: returning -Inf

5: In max(ids, na.rm=TRUE):

no non-missing arguments to max: returning -Inf

6: In max(ids, na.rm = TRUE):

no non-missing arguments to max: returning -Inf

```
> ggplot (df, aes (x= Sepal.Length, y =Petal.Length, color =Species)) +
```

```
geom_point ()+
```

```
labs (title "Scatterplot of Sepal Length and Petal Length grouped by Species", "Sepal  
Length", "Petal Length")
```

- II. Draw a suitable plot to find the skewness of the data for Sepal.Width and print the comment about skewness.

```
> library(dplyr)
```

```
> library(ggplot2)
```

```
> df <- data.frame( Species =c("Setosa", "Versicolour", "Virginia"),
```

```
Sepal.Length =c(5.006, 5.936, 6.500),
```

```
Sepal.Width =c(3.428, 2.77, 2.974), Petal.Length =c(1.462, 4.25, 5.552),
```

```
Petal.Width =c(0.246, 1.326, 2.026)
```

```
> df <- summarise(mean = mean(Sepal.Width))
```

```
group_by (Species)
```

```
summarise (mean = mean(Sepal.Width))
```

```
> ggplot (df_mean, aes(x =Species, y =Sepal.Width)) +
```

```
geom_bar (stat "identity")
```

```
labs (title="Mean Sepal Width based on Species group", x "Species", y "Mean Sepal  
Width")
```

```
> ggplot (df, aes(x =Species, y =Sepal.Width)) +
```

```
geom_boxplot()
```

```
labs (title "Statistical parameter of Sepal Width based on Species group", x "Species",  
y "Sepal Width")
```

```
> ggplot (df, aes (x Sepal.Width, fill Species))+
```

```
geom density (alpha 0.5)+
```

```
labs (title "Density plot of Sepal Width", "Sepal Width")
```

Warning messages:

1: Groups with fewer than two data points have been dropped.

2: Groups with fewer than two data points have been dropped.

3: Groups with fewer than two data points have been dropped.

4: In max(ids, na.rm TRUE):

no non-missing arguments to max: returning -Inf

5: In max(ids, na.rm=TRUE):

no non-missing arguments to max: returning -Inf

6: In max(ids, na.rm=TRUE):

no non-missing arguments to max: returning -Inf

```
> ggplot (df, aes (x Sepal.Length, y Petal.Length, color Species)) +
```

```
geom_point ()+
```

```
labs (title "Scatterplot of Sepal Length and Petal Length grouped by Species", "Sepal  
Length", "Petal Length")
```

- III. Draw ggplot2 scatterplot showing the variables Sepal.Length and Petal.Length grouped by the three-level factor "Species".

```
> library(datasets) > data (iris)
```

```
> head(iris) Sepal.Length
```

```
Sepal.Width 3.5
```

```
Petal.Length
```

```
Petal.Width
```

```
Species 0.2 setosa
```

5.1 4.9

1.4

1.4

0.2

setosa

3.0

3.2 3.1

3

4.7

1.3

0.2

setosa

4.6

1.5

0.2

setosa

5.0

3.6

1.4

0.2

setosa

5.4

3.9

1.7



0.4

setosa

Petal.Width Min. :0.100

Species

> summary(iris) Sepal.Length

Min. 14.300 1st Qu.:5.100

Median :5.800 Mean

setosa :50

Sepal.Width

Min. :2.000 1st

Qu.:2.800

Median :3.000 :3.057

Mean

Petal.Length

Min. 1st Qu.:1.600

:1.000

Median :4.350 Mean

1st Qu.:0.300

Median :1.300

versicolor:50 virginica :50

:5.843

:3.758

Mean

1.199

3rd Qu.:1.800

3rd Qu.:6.400

Max.

:7.900

3rd Qu.:3.300

Max. :4.400

3rd Qu.:5.100

Max.

:6.900

Max.

:2.500