ITA0448-STATICIES WITH R PROGRAMMING

BARATH.P

192121147

ASSESSEMENT-2 DAY-4

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?

PROGRAM:

```
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)
```

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.

PROGRAM:

```
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)

min_age <- min(age)
max_age <- max(age)
print(paste0("Minimum age: ", min_age)) # "Minimum age: 13"
print(paste0("Maximum age: ", max_age)) # "Maximum age: 70"

mean_age <- mean(age)
median_age <- median(age)
print(paste0("Mean age: ", mean_age)) # "Mean age: 28.77778"
print(paste0("Median age: ", median_age)) # "Median age: 25"
hist(age, breaks = seq(10, 80, by = 5), main = "Age Distribution", xlab = "Age")
```

OUTPUT:

```
> 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)
>
> min_age <- min(age)
> max_age <- max(age)
```

```
> print(paste0("Minimum age: ", min_age)) # "Minimum age: 13"
[1] "Minimum age: 13"
> print(paste0("Maximum age: ", max_age)) # "Maximum age: 70"
[1] "Maximum age: 70"
>
> mean_age <- mean(age)
> median_age <- median(age)
> print(paste0("Mean age: ", mean_age)) # "Mean age: 28.77778"
[1] "Mean age: 29.962962962963"
> print(paste0("Median age: ", median_age)) # "Median age: 25"
[1] "Median age: 25"
> hist(age, breaks = seq(10, 80, by = 5), main = "Age Distribution", xla
b = "Age")
>
```

4.Find the categorical column data and convert that to factor form, also find the number of

rows for each factors in dataset.

PROGRAM:

```
data(iris)

dim(iris)
```

OUTPUT:

```
[1] 150   5
```

```
summary(iris)
```

OUTPUT:

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

Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50

1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50

Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50

Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199

3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800

Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500

6.Draw a suitable plot which summaries statistical parameter of Sepal.Width based on

Species group

PROGRAM:

```
# Load the iris dataset
data(iris)

# Create a boxplot of Sepal.Width grouped by Species
boxplot(Sepal.Width ~ Species, data = iris,
        main = "Boxplot of Sepal.Width by Species",
        xlab = "Species", ylab = "Sepal.Width")
```

OUTPUT:

```
> # Load the iris dataset
> data(iris)
>
> # Create a boxplot of Sepal.Width grouped by Species
> boxplot(Sepal.Width ~ Species, data = iris,
+         main = "Boxplot of Sepal.Width by Species",
+         xlab = "Species", ylab = "Sepal.Width")
>

>
```

7. Draw a suitable plot to find the skewness of the data for Sepal.Width and print the

comment about skewness.

PROGRAM:

```
# Load the iris dataset
data(iris)

# Create a histogram of Sepal.Width
hist(iris$Sepal.Width, main = "Histogram of Sepal.Width")

# Create a density plot of Sepal.Width
plot(density(iris$Sepal.Width), main = "Density Plot of
Sepal.Width")
# Load the e1071 package
```

```
library(e1071)


# Calculate the skewness of Sepal.Width

skewness(iris$Sepal.Width)
```

OUTPUT:

```
> # Load the iris dataset
> data(iris)
>
> # Create a histogram of Sepal.Width
> hist(iris$Sepal.Width, main = "Histogram of Sepal.Width")
>
> # Create a density plot of Sepal.Width
> plot(density(iris$Sepal.Width), main = "Density Plot of Sepal.Width
")
> # Load the e1071 package
> library(e1071)
```

8.Draw ggplot2 scatterplot showing the variables Sepal.Length and Petal.Length grouped by

the three-level factor "Species".

PROGRAM:

```
# Load the ggplot2 library

library(ggplot2)
```

```
# Create a scatterplot of Sepal.Length and Petal.Length, colored by Species
ggplot(iris, aes(x = Sepal.Length, y = Petal.Length, color = Species)) +
  geom_point() +
  labs(x = "Sepal.Length", y = "Petal.Length", color = "Species") +
  ggtitle("Relationship between Sepal.Length and Petal.Length by Species")
```

OUTPUT:

```
> # Load the ggplot2 library
> library(ggplot2)
```