**DAY-1 LAB EXPERIMENTS**

**R PROGRAMMING**

**EXPERIMENT 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.**

**(a) What is the mean of the data? What is the median?**

**(b) What is the mode of the data? Comment on the data’s modality (i.e., bimodal, trimodal, etc.).**

**(c) What is the midrange of the data?**

**(d) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?**

**AIM:**

To find the mean and mode of the given data

**MATERIALS REQUIRED:**

R Software

**PROGRAM:**

# Given data

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)

# Mean

mean\_age <- mean(age)

cat("Mean:", mean\_age, "\n")

# Median

median\_age <- median(age)

cat("Median:", median\_age, "\n")

# Mode

mode\_age <- as.numeric(names(sort(-table(age))[1]))

cat("Mode:", mode\_age, "\n")

# Modality

modality <- length(unique(age))

cat("Modality:", modality, "-modal\n")

# Midrange

midrange\_age <- (min(age) + max(age)) / 2

cat("Midrange:", midrange\_age, "\n")

# Quartiles

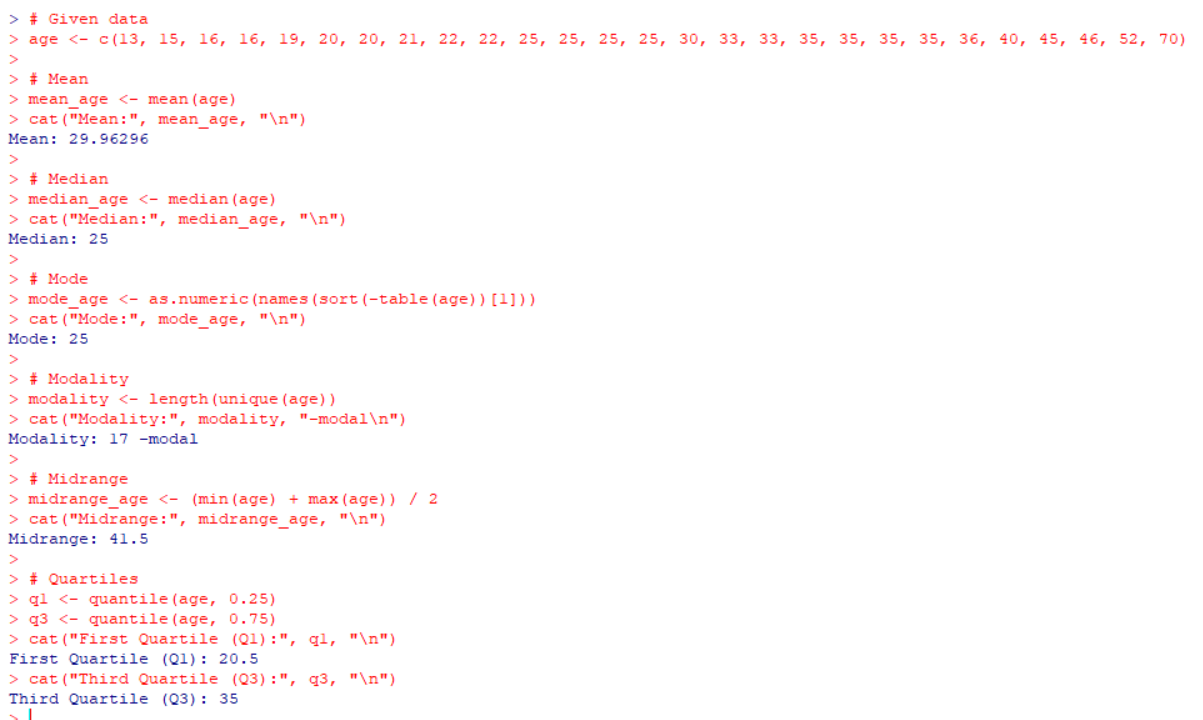
q1 <- quantile(age, 0.25)

q3 <- quantile(age, 0.75)

cat("First Quartile (Q1):", q1, "\n")

cat("Third Quartile (Q3):", q3, "\n")

**OUTPUT:**

****

**EXPERIMENT 2:**

**.Data Preprocessing :Reduction and Transformation**

**Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000 (a) min-max normalization by setting min = 0 and max = 1 (b) z-score normalization**

**AIM:**

To find the min max and z score normalisation of the given data

**MATERIALS REQUIRED:**

R Software

**PROGRAM:**

# Create the array

data1 <- c(200, 300, 400, 600, 1000)

# Min-max normalization

min\_max <- (data1 - min(data1)) / (max(data1) - min(data1))

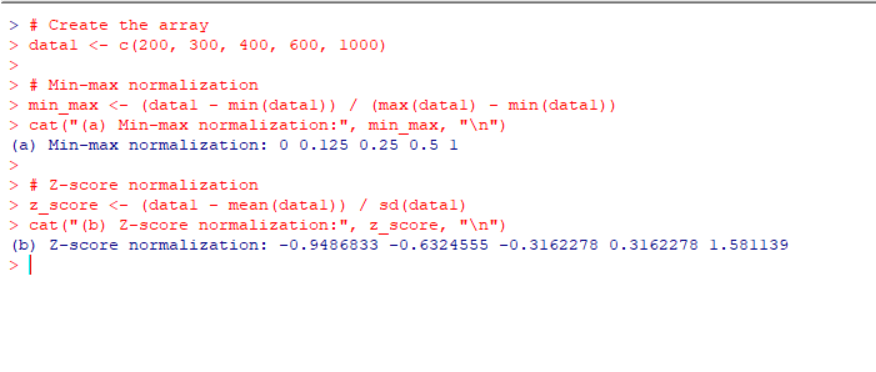
cat("(a) Min-max normalization:", min\_max, "\n")

# Z-score normalization

z\_score <- (data1 - mean(data1)) / sd(data1)

cat("(b) Z-score normalization:", z\_score, "\n")

**OUTPUT:**

****

**EXPERIMENT 3**

**Data:11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,**

**72,73,75**

**a) Smoothing by bin mean**

**b) Smoothing by bin median**

**c) Smoothing by bin boundaries**

**AIM:**

To find the mean,median and boundaries of the given data by using smoothing the bin.

**MATERIALS REQUIRED:**

R Software

**PROGRAM:**

data2 <- c(11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75)

# Smoothing by bin mean

smooth\_mean <- tapply(data2, ceiling(seq\_along(data2)/5), mean)

# Smoothing by bin median

smooth\_median <- tapply(data2, ceiling(seq\_along(data2)/5), median)

# Smoothing by bin boundaries

smooth\_boundaries <- tapply(data2, ceiling(seq\_along(data2)/5), function(x) c(min(x), max(x)))

# Combine all results into a data frame

result\_df <- data.frame(Bin = 1:length(smooth\_mean),

Bin\_Boundaries = unlist(lapply(smooth\_boundaries, paste, collapse="-")),

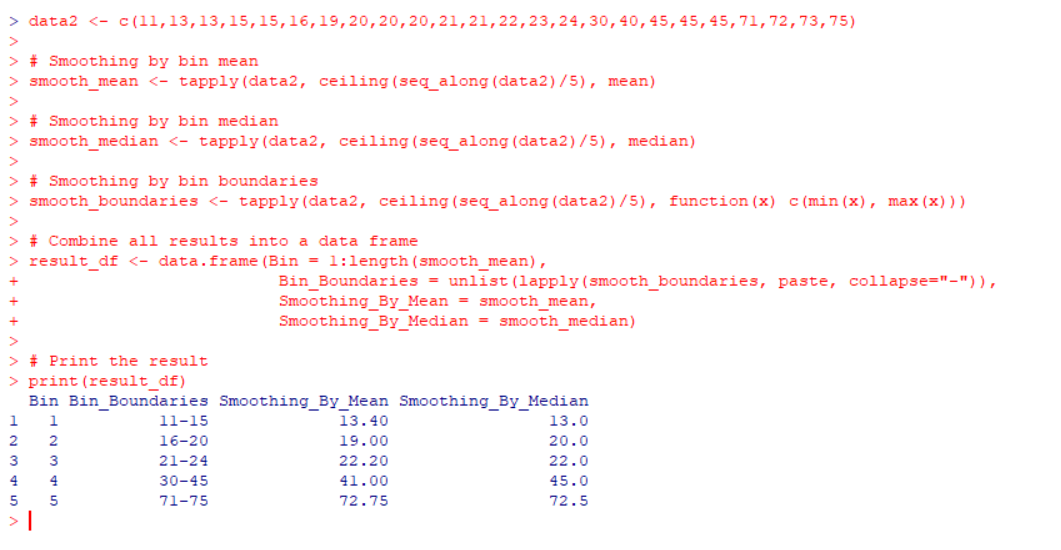
Smoothing\_By\_Mean = smooth\_mean,

Smoothing\_By\_Median = smooth\_median)

# Print the result

print(result\_df)

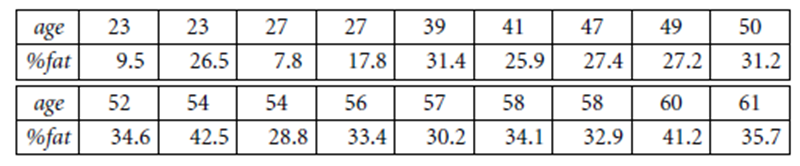
**OUTPUT:**



**EXPERIMENT 4:**

**Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:**

**(a) Calculate the mean, median, and standard deviation of age and %fat. (b) Draw the boxplots for age and %fat.**

**(c) Draw a scatter plot and a q-q plot based on these two variables**

**AIM:**

To find the mean,median,standard deviation of the given data

**MATERIALS REQUIRED:**

R Software

**PROGRAM:**

**age <- c(23, 25, 32, 35, 47, 48, 52, 54, 54, 56, 57, 58, 58, 60, 61, 64, 65, 68)**

**body\_fat <- c(9.5, 26.5, 7.8, 17.8, 31.4, 25.9, 27.4, 31.2, 34.6, 42.5, 28.8, 33.4, 30.2, 34.1, 32.9, 41.2, 35.7, 30.2)**

**# (a) Calculate mean, median, and standard deviation**

**mean\_age <- mean(age)**

**median\_age <- median(age)**

**sd\_age <- sd(age)**

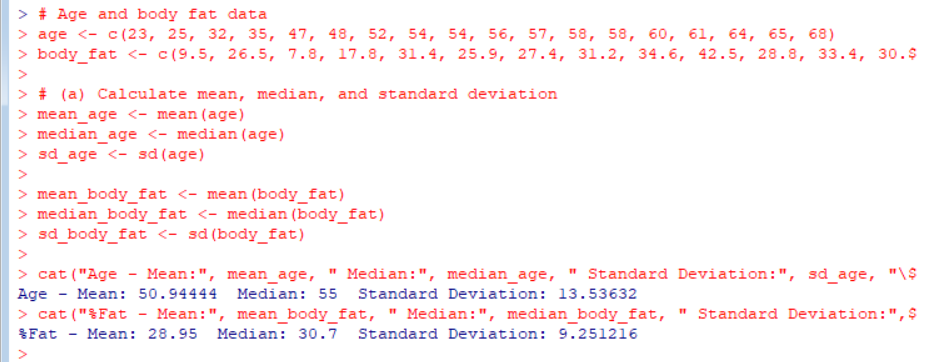
**mean\_body\_fat <- mean(body\_fat)**

**median\_body\_fat <- median(body\_fat)**

**sd\_body\_fat <- sd(body\_fat)**

**cat("Age - Mean:", mean\_age, " Median:", median\_age, " Standard Deviation:", sd\_age, "\n")**

**cat("%Fat - Mean:", mean\_body\_fat, " Median:", median\_body\_fat, " Standard Deviation:", sd\_body\_fat, "\n")**

****

**# Age and body fat data**

**age <- c(23, 25, 32, 35, 47, 48, 52, 54, 54, 56, 57, 58, 58, 60, 61, 64, 65, 68)**

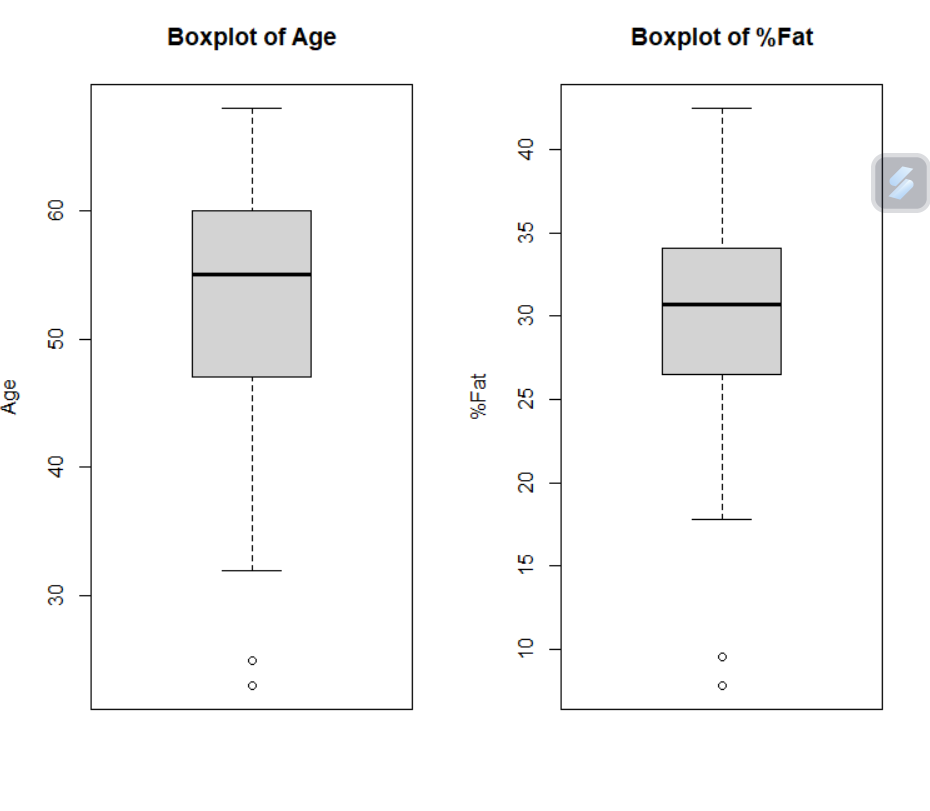
**body\_fat <- c(9.5, 26.5, 7.8, 17.8, 31.4, 25.9, 27.4, 31.2, 34.6, 42.5, 28.8, 33.4, 30.2, 34.1, 32.9, 41.2, 35.7, 30.2)**

**# Draw boxplots**

**par(mfrow=c(1,2)) # Setting up the plot layout as 1x2 grid**

**boxplot(age, main="Boxplot of Age", ylab="Age")**

**boxplot(body\_fat, main="Boxplot of %Fat", ylab="%Fat")**

****

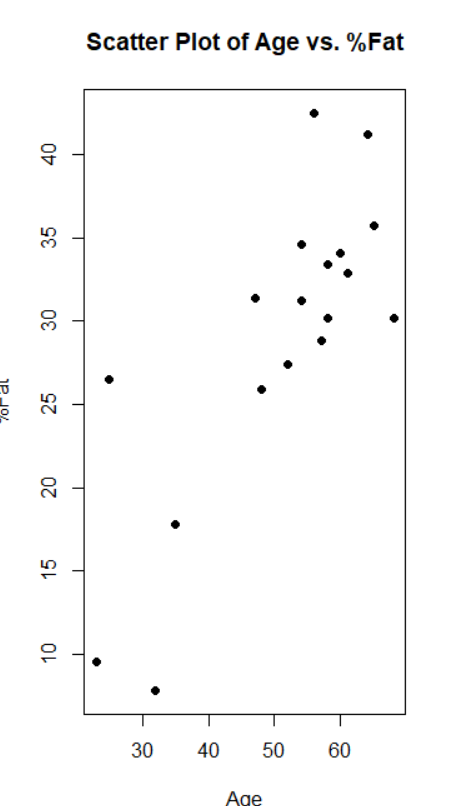
**# Age and body fat data**

**age <- c(23, 25, 32, 35, 47, 48, 52, 54, 54, 56, 57, 58, 58, 60, 61, 64, 65, 68)**

**body\_fat <- c(9.5, 26.5, 7.8, 17.8, 31.4, 25.9, 27.4, 31.2, 34.6, 42.5, 28.8, 33.4, 30.2, 34.1, 32.9, 41.2, 35.7, 30.2)**

**# Scatter plot**

**plot(age, body\_fat, main="Scatter Plot of Age vs. %Fat", xlab="Age", ylab="%Fat", pch=19)**

****

**# Age and body fat data**

**age <- c(23, 25, 32, 35, 47, 48, 52, 54, 54, 56, 57, 58, 58, 60, 61, 64, 65, 68)**

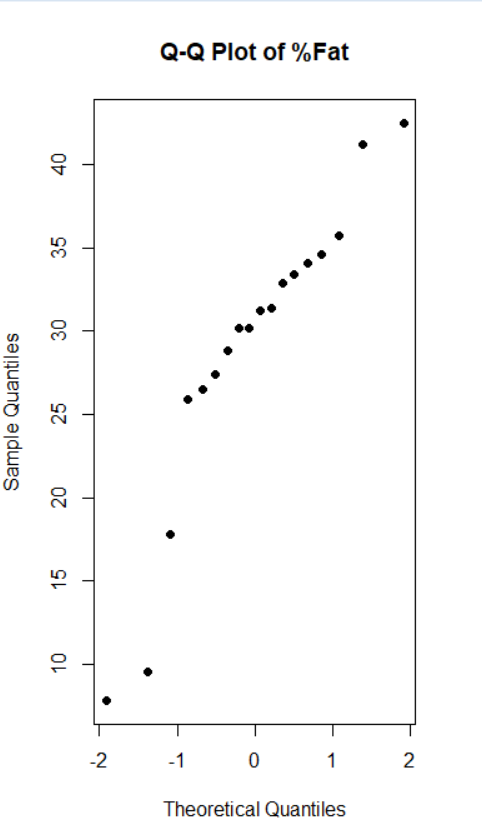
**body\_fat <- c(9.5, 26.5, 7.8, 17.8, 31.4, 25.9, 27.4, 31.2, 34.6, 42.5, 28.8, 33.4, 30.2, 34.1, 32.9, 41.2, 35.7, 30.2)**

**# Q-Q plot for age**

**qqplot(qnorm(ppoints(length(age))), sort(age), main="Q-Q Plot of Age", xlab="Theoretical Quantiles", ylab="Sample Quantiles", pch=19)**

**# Q-Q plot for body fat**

**qqplot(qnorm(ppoints(length(body\_fat))), sort(body\_fat), main="Q-Q Plot of %Fat", xlab="Theoretical Quantiles", ylab="Sample Quantiles", pch=19)**

****

**EXPERIMENT 5:**

**.Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:**

**(i) Use min-max normalization to transform the value 35 for age onto the range [0.0, 1.0].  
 (ii) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.  
 (iii) Use normalization by decimal scaling to transform the value 35 for age. Perform the above functions using R – tool**

**AIM:**

To find the minmax,z score,decimal scaling of the given data

**MATERIALS REQUIRED:**

R Software

**PROGRAM:**

# Value to normalize

age <- 35

# Standard deviation for z-score normalization

std\_dev <- 12.94

# Min-max normalization

min\_max <- function(x) {

(x - min(x)) / (max(x) - min(x))

}

# Z-score normalization

z\_score <- function(x, mean, std\_dev) {

(x - mean) / std\_dev

}

# Decimal scaling normalization

decimal\_scaling <- function(x) {

max\_log <- floor(log10(max(x)))

x / (10^max\_log)

}

# Performing normalization

# Min-max normalization

min\_max\_result <- min\_max(age)

# Z-score normalization

z\_score\_result <- z\_score(age, mean = 0, std\_dev = std\_dev)

# Decimal scaling normalization

decimal\_scaling\_result <- decimal\_scaling(age)

# Printing results

cat("Min-max normalization result:", min\_max\_result, "\n")

cat("Z-score normalization result:", z\_score\_result, "\n")

cat("Decimal scaling normalization result:", decimal\_scaling\_result, "\n")