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:

```
> # 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")
Mean: 29.96296
>
> # Median
> median_age <- median(age)
> cat("Median:", median_age, "\n")
Median: 25
>
> # Mode
> mode_age <- as.numeric(names(sort(-table(age))[1]))
> cat("Mode:", mode_age, "\n")
Mode: 25
>
> # Modality
> modality <- length(unique(age))
> cat("Modality:", modality, "-modal\n")
Modality: 17 -modal
>
> # Midrange
> midrange_age <- (min(age) + max(age)) / 2
> cat("Midrange:", midrange_age, "\n")
Midrange: 41.5
>
> # Quartiles
> q1 <- quantile(age, 0.25)
> q3 <- quantile(age, 0.75)
> cat("First Quartile (Q1):", q1, "\n")
First Quartile (Q1): 20.5
> cat("Third Quartile (Q3):", q3, "\n")
Third Quartile (Q3): 35
```

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:

```
> # 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")
(a) Min-max normalization: 0 0.125 0.25 0.5 1
>
> # Z-score normalization
> z_score <- (data1 - mean(data1)) / sd(data1)
> cat("(b) Z-score normalization:", z_score, "\n")
(b) Z-score normalization: -0.9486833 -0.6324555 -0.3162278 0.3162278 1.581139
>
```

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:

```
> 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)
  Bin Bin_Boundaries Smoothing_By_Mean Smoothing_By_Median
1   1          11-15             13.40                13.0
2   2          16-20             19.00                20.0
3   3          21-24             22.20                22.0
4   4          30-45             41.00                45.0
5   5          71-75             72.75                72.5
>
```

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

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

AIM:

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

MATERIALS REQUIRED:

R Software

PROGRAM:

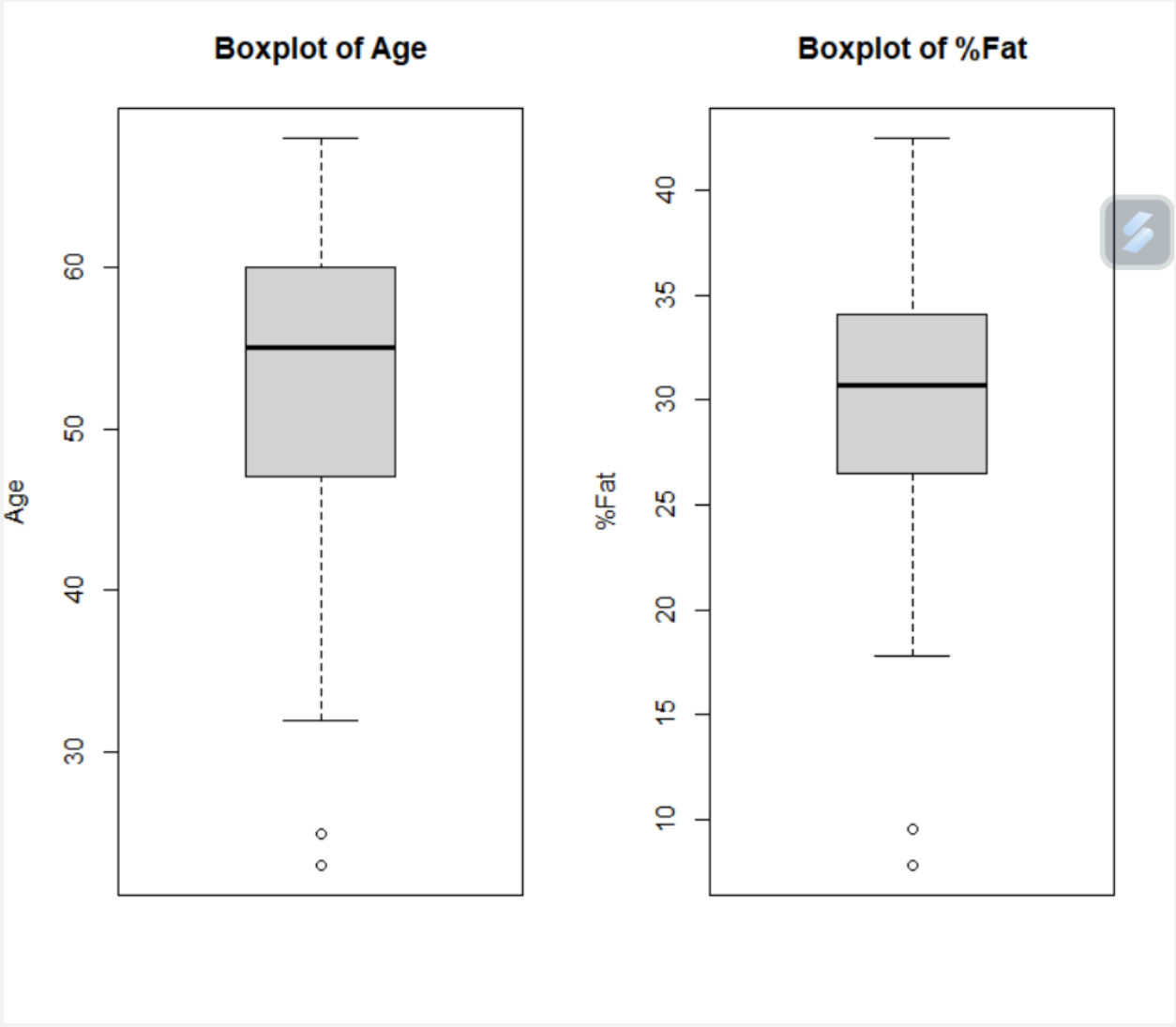
```
age <- c(23, 25, 32, 35, 47, 48, 52, 54, 54, 56, 57, 58, 58, 60, 61, 64, 65, 68)
```

```
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.$
>
> # (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, "\$
Age - Mean: 50.94444  Median: 55  Standard Deviation: 13.53632
> cat("%Fat - Mean:", mean_body_fat, " Median:", median_body_fat, " Standard Deviation:",$
%Fat - Mean: 28.95  Median: 30.7  Standard Deviation: 9.251216
>
```

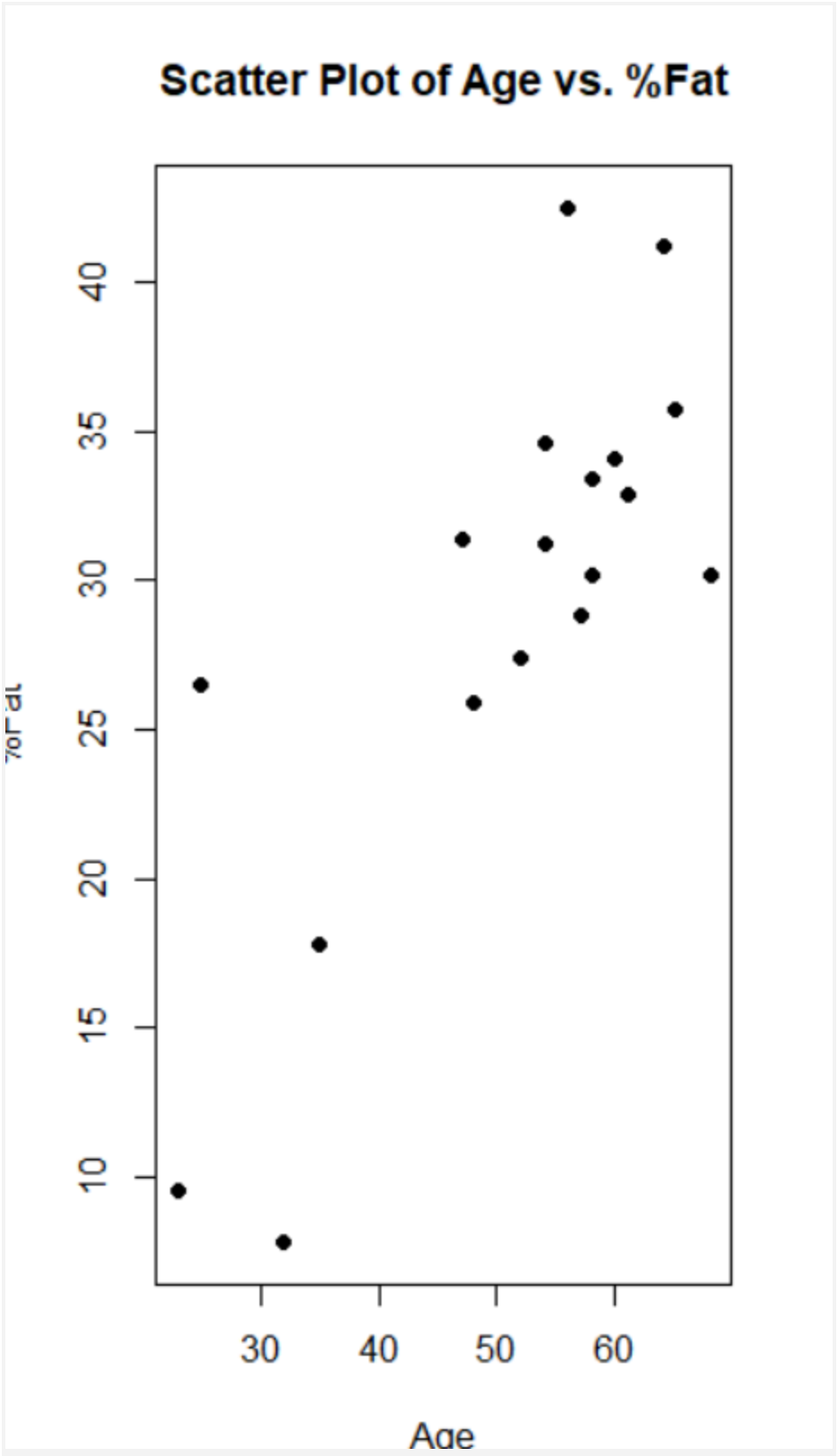
```
# Age and body fat data

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")
```



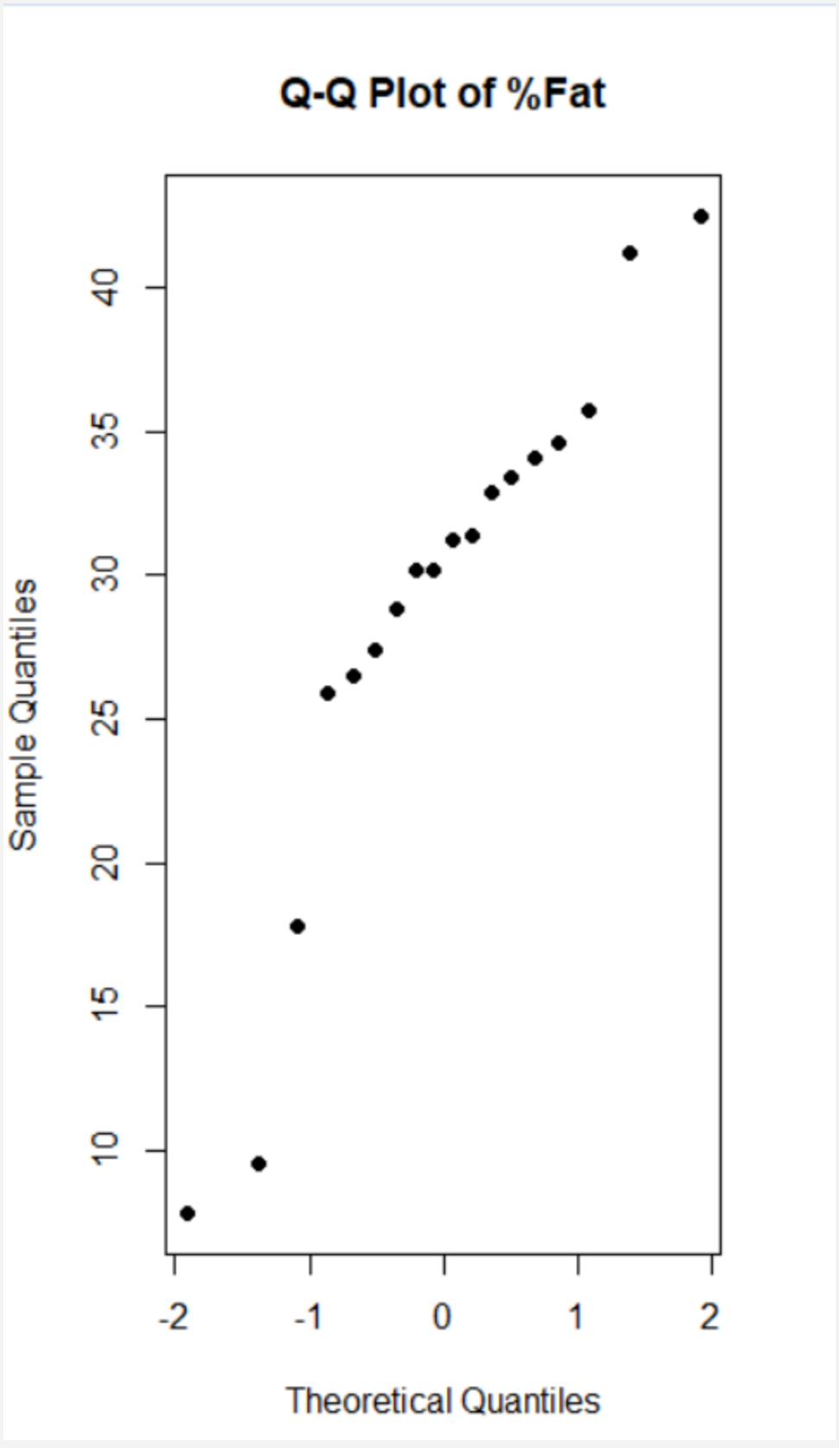

```
# Age and body fat data
```

```
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)
```


Q-Q Plot of %Fat
Sample Quantiles
10
15
20
25
30
35
40
-2
-1
0
1
2
Theoretical Quantiles

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")
```