Part - B

1. Write a R program to create a Vector containing the following 8 values and perform the following operations.

4 3 0 5 2 9 4 5

a. Find mean, median, mode.
b. Find the range.
c. Find the 35th and 78th percentile.
d. Find the variance and standard deviation
e. Find the interquartile range.
f. Find the z-score for each value.

```
values <- c(4, 3, 0, 5, 2, 9, 4, 5)

mean_value <- mean(values)
median_value <- median(values)
table_values <- table(values)
mode_value <- as.numeric(names(table_values)[table_values == max(table_values)])
range_value <- range(values)

percentile_35 <- quantile(values, 0.35)
percentile_78 <- quantile(values, 0.78)

variance_value <- var(values)
std_deviation_value <- sd(values)

IQR_value <- IQR(values)

z_scores <- scale(values)

cat("\nMean:", mean_value)
cat("\nMedian:", median_value)
cat("\nMode:", mode_value)
cat("\nRange:", range_value[2] - range_value[1])
cat("\n35th Percentile:", percentile_35)
cat("\n78th Percentile:", percentile_78)
cat("\nVariance:", variance_value)
cat("\nStandard Deviation:", std_deviation_value)
cat("\nInterquartile Range:", IQR_value)
cat("\nZ-Scores:", z_scores)
```

Output:

```
Mean: 4
Median: 4
Mode: 4 5
```

Range: 9
35th Percentile: 3.45
78th Percentile: 5
Variance: 6.857143
Standard Deviation: 2.618615
Interquartile Range: 2.25
Scores: 0 -0.3818813 -1.527525 0.3818813 -0.7637626 1.909407 0 0.3818813

Manual:
Vector values = (4 3 0 5 2 9 4 5)
Sorted vector values = (0 2 3 4 4 5 5 9)
No. of data in vector = n = 8

a. Find mean, median, mode.

$$Mean = \frac{4+3+0+5+2+9+4+5}{8} = 4$$

$$Median = \frac{[\frac{8}{2}^{th} value + \frac{8}{2}+1^{th} value]}{2}$$

$$= \frac{[4+4]}{2} = \frac{8}{2} = 4$$

Mode = 4, 5

b. Find the range.
Range: max(vector values) − min(vector values) = 9 − 0 = 9

c. Find the 35th and 78th percentile.

$$P_{35} = (((0.35 * (n - 1)) + 1)^{th}$$
$$= (((0.35 * 7) + 1)^{th} = 3.45^{th}$$
$$= 3^{rd} value + (0.45 * (4^{th} value - 3^{rd} value))$$
$$P_{35} = 3 + (0.45 * (4 - 3)) = 3.45$$

$$P_{78} = (((0.78 * (n - 1)) + 1)^{th}$$
$$= (((0.78 * 7) + 1)^{th} = 6.46^{th}$$
$$= 6^{th} value + (0.46 * (7^{th} value - 6^{rd} value))$$
$$P_{78} = 5 + (0.46 * (5 - 5)) = 5$$

d. Find the variance and standard deviation

x	x - μ	$(x - \mu)^2$

4	0	0
3	-1	1
0	-4	16
5	1	1
2	-2	4
9	5	25
4	0	0
5	1	1
Σx = 32		$\Sigma(x - \mu)^2 = 48$

$$\mu = \frac{\Sigma x}{N}$$
$$= \frac{32}{8}$$
$$\mu = 4$$

$$variance,\ \sigma^2 = \frac{\Sigma(x-\mu)^2}{N-1}$$
$$\sigma^2 = \frac{48}{8-1} = \frac{48}{7}$$
$$\sigma^2 = 6.857143$$

$$Standard\ Deviation,\ \sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N-1}}$$
$$\sigma = \sqrt{\frac{48}{7}}$$
$$\sigma = 2.618615$$

e. Find the interquartile range.

$$Q1 = P_{25} = (((0.25 * (n - 1)) + 1)^{th}$$
$$= ((0.25 * 7) + 1)^{th} = 2.75^{th}$$
$$= 2^{nd} value + (0.75 * (3^{rd} value - 2^{nd} value))$$
$$Q1 = P_{25} = 2 + (0.75 * (3 - 2)) = 2.75$$
$$Q3 = P_{75} = (((0.75 * (n - 1)) + 1)^{th}$$
$$= ((0.75 * 7) + 1)^{th} = 6.25^{th}$$
$$= 6^{th} value + (0.25 * (7^{th} value - 6^{rd} value))$$
$$Q3 = P_{75} = 5 + (0.25 * (5 - 5)) = 5$$

$$InterQuartile\ Range, IQR = Q3 - Q1 = 5 - 2.75 = 2.25$$

f. Find the z-score for each value.

$z\ scores,\ z = \frac{x-\mu}{\sigma}$

z scores	x	x - μ	$\frac{x-\mu}{\sigma}$
z(4)	4	0	0.0000000
z(3)	3	-1	-0.3818813
z(0)	0	-4	-1.5275250
z(5)	5	1	0.3818813
z(2)	2	-2	-0.7637625
z(9)	9	5	1.9094063
z(4)	4	0	0.0000000
z(5)	5	1	0.3818813

<u>Note:</u> Only If examiner ask for find or give value in percentile position(rank)
Do this,

In Program code:(not sure it is correct way)
Use this:

```
percentile_35 <- sorted_vector[floor(quantile(values, 0.35))]
percentile_78 <- sorted_vector[floor(quantile(values, 0.78))]
```

In Manual:
Use this:

$$P_{35} = ((0.35 * (n - 1)) + 1)^{th}$$
$$= ((0.35 * 7) + 1)^{th} = 3.45^{th}$$
$$= 3^{rd} value + (0.45 * (4^{th} value - 3^{rd} value))$$
$$P_{35} = 3 + (0.45 * (4 - 3)) = 3.45$$

As 3.45 is in decimal, round it up. i.e 3
Find the 3rd value from the vector. i.e 3

$$P_{78} = ((0.78 * (n - 1)) + 1)^{th}$$
$$= ((0.78 * 7) + 1)^{th} = 6.46^{th}$$
$$= 3^{rd} value + (0.46 * (4^{th} value - 3^{rd} value))$$
$$P_{78} = 5 + (0.46 * (5 - 5)) = 5$$

As 5 is in integer, you don't need to round it up. i.e 5
Find the 3rd value from the vector. i.e 4

2. Write R script to find the correlation coefficient and type of correlation between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method (Direct Method).

Firm	1	2	3	4	5	6	7	8	9	10
Advertisement Exp. (Rs. in Lakhs)	11	13	14	16	16	15	15	14	13	13
Sales Volume (Rs. in Lakhs)	50	50	55	60	65	65	65	60	60	50

```
advertisement_expenses <- c(11, 13, 14, 16, 16, 15, 15, 14, 13, 13)
sales_volume <- c(50, 50, 55, 60, 65, 65, 65, 60, 60, 50)

correlation_coefficient <- cor(advertisement_expenses, sales_volume, method = "pearson")

type_of_correlation_coefficient <- if(correlation_coefficient > 0){
  "Positive correlation"
} else if(correlation_coefficient < 0){
  "Negative correlation"
}else{
  "No correlation"
}
cat("Correlation Coefficient:", correlation_coefficient, "\n")
cat("Type of Correlation:", type_of_correlation_coefficient, "\n")
```

Output:
Correlation Coefficient: 0.7865665
Type of Correlation: Positive correlation

Manual:

Let us assume that advertisement expenses are variable x and sales volume are variable y.

Calculation of Karl Pearson's coefficient of correlation (Direct Method)

Firm	x series		y series		xy
	x	x^2	y	y^2	

1	11	121	50	2500	550
2	13	169	50	2500	650
3	14	196	55	3025	770
4	16	256	60	3600	960
5	16	256	65	4225	1040
6	15	225	65	4225	975
7	15	225	65	4225	975
8	14	196	60	3600	840
9	13	169	60	3600	780
10	13	169	50	2500	650
$n = 10$	$\Sigma x = 140$	$\Sigma x^2 = 1982$	$\Sigma y = 580$	$\Sigma y^2 = 34000$	$\Sigma xy = 8190$

Using Karl Pearson's coefficient of correlation method (Direct Method) formula

$$r = \frac{(n \,.\, \Sigma xy) - (\Sigma x \,.\, \Sigma y)}{(\sqrt{(n \,.\, \Sigma x^2) - (\Sigma x)^2}) \,.\, (\sqrt{(n \,.\, \Sigma y^2) - (\Sigma y)^2})}$$

$$r = \frac{(10 \,.\, 8190) - (140 \,.\, 580)}{(\sqrt{(10 \,.\, 1982) - (140)^2}) \,.\, (\sqrt{(10 \,.\, 34000) - (580)^2})}$$

$$r = \frac{81900 - 81200}{(\sqrt{19820 - 19600}\ ) \,.\, (\sqrt{340000 - 336400})}$$

$$r = \frac{700}{(\sqrt{220}) \,.\, (\sqrt{3600})}$$

$$r = \frac{700}{14.832397 \,.\, 60}$$

$$r = \frac{700}{889.94376} = 0.78657$$

Interpretation: From the above calculation it is very clear that there is a high degree of positive correlation i.e. r = 0.78657 between the two variables.
i.e Increase in advertisement expenses leads to increased sales volume.