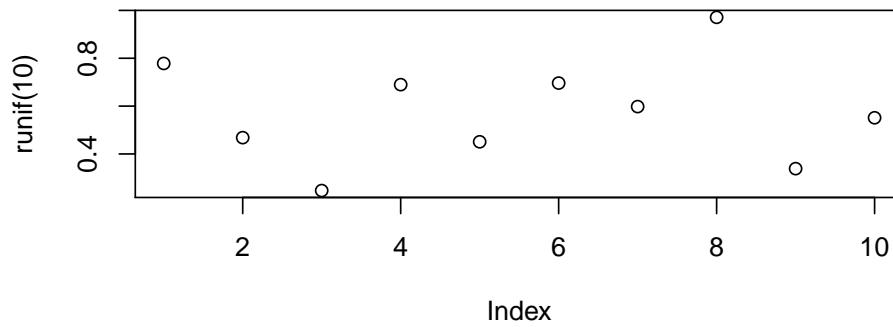1. Consider the below outputs generated in R.

(a) (5 Points) Please write the down the R command that will provide the below plot. Describe in detail what the points in the plot represent.

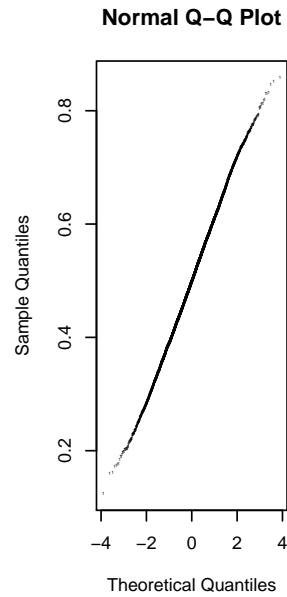(b) (5 Points) Suppose the below commands are entered in R as given. Please fill in the blanks appropriately:
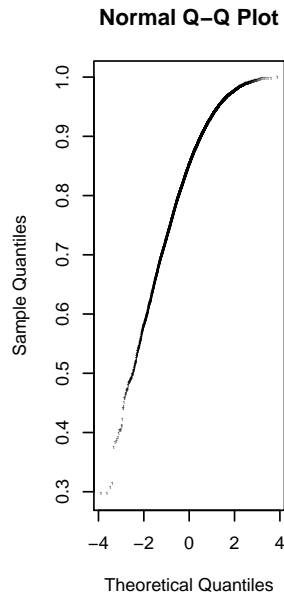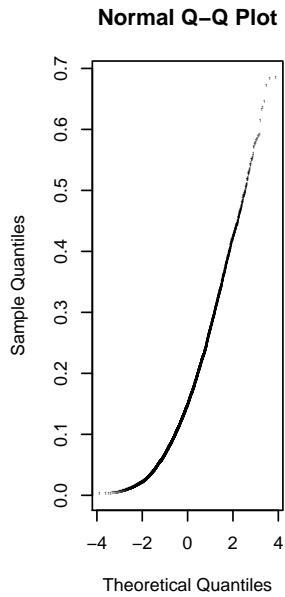
```
> x = c (7,-12,9,15,NA,-8,14,NA)
> x
```

⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺

```
> which(is.na(x))
```

⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺

```
> sum(!is.na(x))
```

⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺

```
> x >0
```

⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺⸺

(c) (5 Points) Describe the plots below. With suitable justification decide (if possible) whether the data distribution is skewed Left, Right or Symmetric.

2. Sampagni Car company has made its latest COOL-X above road surface car. Below is a scatter plot of Distance taken to Stop versus Speed of a car taken 50 times.
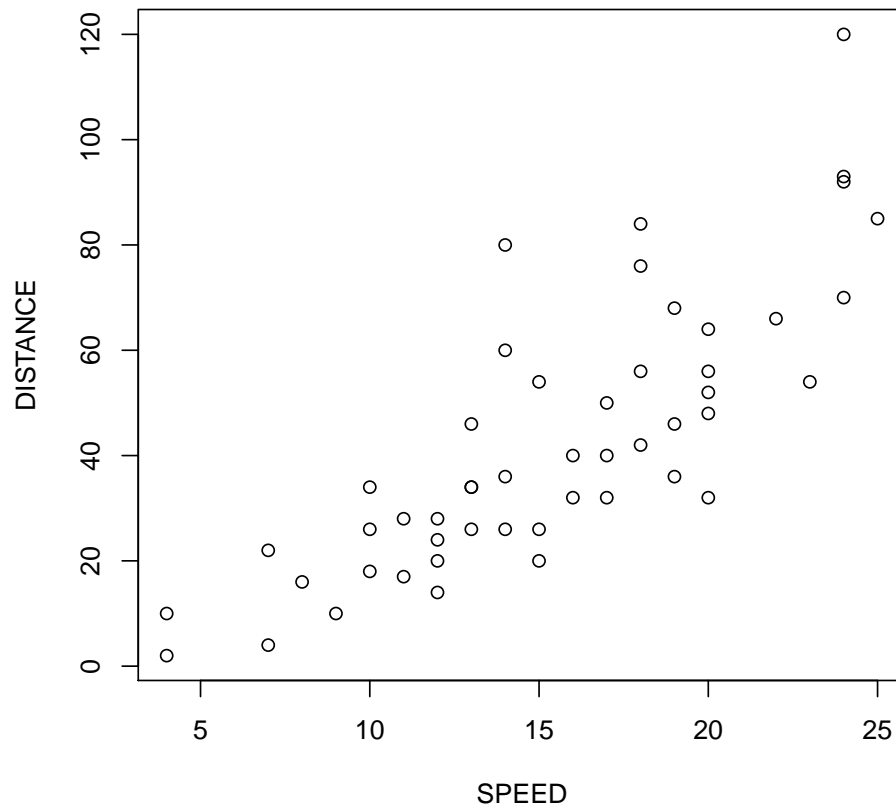
(a) (5 points) Using the output of R below:

```
> lm(DISTANCE~SPEED)

Call:
lm(formula = DISTANCE ~ SPEED)

Coefficients:
(Intercept)        SPEED
   -17.579        3.932
```

on scatter-plot draw the best possible linear relationship for the above data.

(b) (8 points) Suppose we denote the 50 data points as

$$\{(y_i, x_i) : 1 \leq i \leq 50\}$$

Describe the output of the R command in part(a) in terms of the data points.

(c) (2 points) Predict the distance taken to stop when the speed is 10000

3.(a) (6 points) B.Math(hons.) second year class are trying to determine the ideal amount of coffee to be sold at the canteen. They randomly sample 8 students in the canteen and ask them how much *ml* of coffee they would like to drink. The data is shown below.

$$300, 500, 300, 700, 300, 500, 300, 700.$$

Assuming normality and variance to be 100, construct a 95% confidence interval for the true population mean of the preferred drinking amount using the above data.

(b) (4 points) Below is an R function intending to find the 95% confidence interval for true mean from data x with known variance 1. Please fill in the blanks.

```
cifn = function(x) {

z = qnorm( _____ )

sdx = _____

c(mean(x) - z*sdx, mean(x) + z*sdx)

}
```

(c) (5 points) If `cifn` is above, then please explain the intention of the below program and the output given.

```
> normaldata = replicate(100, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)


TRUEIN
FALSE  TRUE
    4    96
```

4. Siva's class in 1993 of 15 received the following scores out of 50:

$$21, 22, 25, 28, 29, 30, 32, 33, 33, 34, 36, 40, 41, 41, 43.$$

(a) (5 points) Find the five number summary for the data.

(b) (2 points) What is the intraquartile range of the data ?

(c) (2 points) Are there any outliers in the dataset ? Explain.

(d) (4 points) Draw a boxplot for the data.

(e) (2 points) Describe if the plot is (best option amongst) left skewed, right skewed, symmetric.