**ASSIGNMENT-3**

**8.2) a**

SAMPLE MEAN:

=

= 17.954

**VARIANCE**:

S2

S2

= 16605.75 – (50\*17.954\*17.954) / 50 – 1

= 9.968

**STANDARD DEVIATION**:

S =

S =

= 3.157

b) **STANDARD ERROR OF THE SAMPLE MEAN**:

s() = s /

s() = 3.157 /

s() = 0.4464

c) **FIVE POINT SUMMARY:**

five-point summary = ( min Xi, , , , max Xi )

MEDIAN =

Here, n is 50

=

= 17.55

min Xi = 11.9

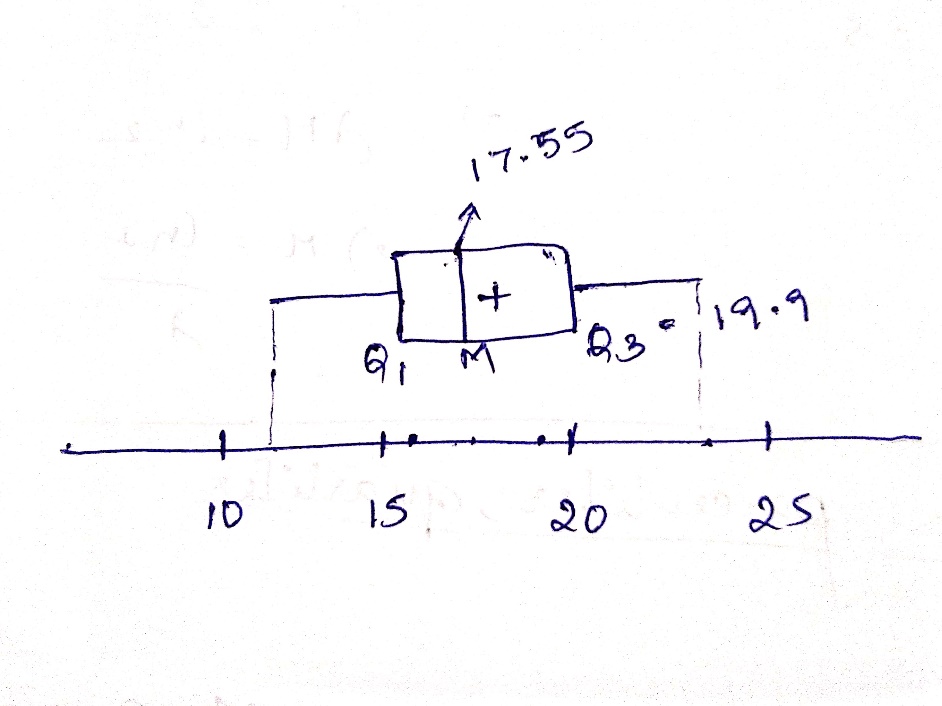
max Xi = 24.1

= 15.8 (np = 50 \* 0.25 = 12.5, no more than 12.5 observations, 13th observation)

= 19.9 (38th observation)

five-point summary = (11.9, 15.8, , , 24.1)

**BOX PLOT:**

****

**d) = −** .

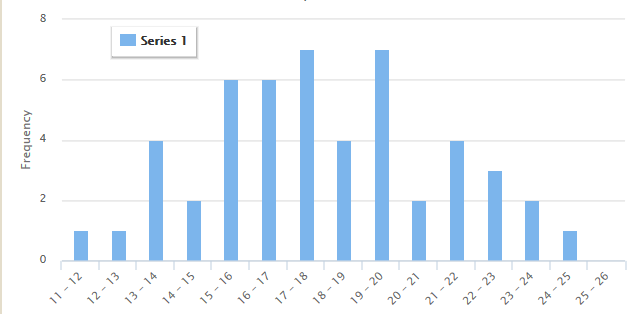
= 19.9 – 15.8

= 4.1

**– 1.5() = 15.8 – 1.5(4.1) = 9.65**

**+ 1.5() = 19.9 + 1.5(4.1) = 26.05**

There are no outliers.

e) 

No. Number of concurrent users do not follow Normal distribution.

8.4)

Q1 = 0.25

The Z-score value for Q1 is -0.674

Q3 = 0.75

The Z-score value for Q3 is 0.674

IQR = Q3 – Q1 = 0.674 – (-0.674)

= 1.348

Probability of random variable within 1.5IQR

P(Q1 – 1.5IQR < z < Q3 + 1.5IQR)

P(-2.696 < Z < 2.696)

P(2.696) – P(-2.696)

0.9964 – 0.0036

0.993

8.9) a)

Given data: 43, 37, 50, 51, 58, 105, 52, 45, 45, 10.

Sorted data: 10, 37, 43, 45, 45, 50, 51, 52, 58, 105

n = 10

1. **Mean:**

= 496 / 10

= 49.6

**MEDIAN** =

= (45 + 50 ) / 2

= 47.5

**QUARTILES** = Q1 = 43 (np = 2.5, so the Q1 should not be more than 2 observations)

Q3 = 52 (n(1-p) = 7.5, So the next smallest observation is 8)

Q2 = 47.5

**STANDARD DEVIATION**:

S =

S2

S2

S2=

S2= 551.155

S =

S = 23.476

1. **= −** .

= 52– 43

= 9

**– 1.5() = 43 – 1.5(9) = 29.5**

**+ 1.5() = 52 + 1.5(9) = 65.5**

The outlier to the left is 10

The outlier to the right is 105

1. MEAN without outliers: 47.625

MEDIAN without outliers =

= (45+50) / 2

= 47.5

QUARTILES without outliers= Q1 = 44 (np = 2, (43 + 45)/2)

Q3 = 51.5 (n(1-p) = 6, (51 + 52)/2)

Q2 = 47.5

STANDARD DEVIATION without outliers:

S =

S2

S2

S2=

S2= 41.69

S =

S = 6.456

1. An outlier is a value which is very different from the rest of the data in the dataset.

As mean is calculated by taking each dataset into consideration, it is greatly affected by the outliers. Outliers that are much higher than the mid point of the data, inflate the mean, whereas the outliers with less values decreases the mean significantly. Median is least effected by the outliers. Interquartile ranges are used to determine the outliers. Standard deviation and variance are affected by the outliers.