

Assignment No.1

- 1) Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows:

age	frequency
1–5	200
6–15	450
16–20	300
21–50	1500
51–80	700
81–110	44

Compute an approximate median value for the data.

**Answer:**

Age Interval	Frequency	Cumulative Frequency
1–5	200	200
6–15	450	650
16–20	300	950
21–50	1500	2450
51–80	700	3150
81–110	44	3194

**total frequency ( $n$ )** = 200+450+300+1500+700+44=3194

The formula for the approximate median is:

$$\text{Median} = L + \left( \frac{\frac{n}{2} - F}{f} \right) \times w$$

- $L$  = lower boundary of the median interval = 21
- $n$  = total frequency = 3194
- $F$  = cumulative frequency before the median interval = 950
- $f$  = frequency of the median interval = 1500
- $w$  = width of the median interval =  $50 - 21 + 1 = 30$

$$\text{Median} = 21 + \left( \frac{1597 - 950}{1500} \right) \times 30$$

$$\text{Median} = 21 + \left( \frac{647}{1500} \right) \times 30$$

$$\text{Median} = 21 + 12.94$$

$$\text{Median} \approx 33.94$$

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

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

(a) Calculate the mean, median, and standard deviation of age and %fat

(b) Plot the Boxplot of age and %fat

**Answer :**

$$\text{Mean (Age)} = \frac{\sum \text{Age}}{N} = \frac{836}{18} = 46.5$$

$$\text{Median (Age)} = \frac{50 + 52}{2} = 51$$

$$\text{Standard Deviation} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Where:

- $x_i$  = each data point
- $\bar{x}$  = mean (46.5)
- $n$  = number of data points (18)

Age ( $x_i$ )	Mean ( $\bar{x}$ )	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
23	46.4	-23.4	547.56
23	46.4	-23.4	547.56
27	46.4	-19.4	376.36
27	46.4	-19.4	376.36
39	46.4	-7.4	54.76
41	46.4	-5.4	29.16
47	46.4	0.6	0.36
49	46.4	2.6	6.76
50	46.4	3.6	12.96
52	46.4	5.6	31.36
54	46.4	7.6	57.76
54	46.4	7.6	57.76
56	46.4	9.6	92.16
57	46.4	10.6	112.36
58	46.4	11.6	134.56
58	46.4	11.6	134.56

Age ( $x_i$ )	Mean ( $\bar{x}$ )	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
---------------	--------------------	-----------------	---------------------

60	46.4	13.6	184.96
----	------	------	--------

61	46.4	14.6	213.16
----	------	------	--------

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{2970.12}{18} = 165.01$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}} = \sqrt{165.01} = 12.85$$

**%fat**

$$\text{Mean (\%fat)} = 498 / 18 = 27.63$$

$$\text{Median (\%fat)} = 30.7$$

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
-------	-----------------	---------------------

9.5	-17.07	291.38
-----	--------	--------

26.5	-0.07	0.00
------	-------	------

7.8	-18.77	352.31
-----	--------	--------

17.8	-8.77	76.91
------	-------	-------

31.4	4.83	23.33
------	------	-------

25.9	-0.67	0.45
------	-------	------

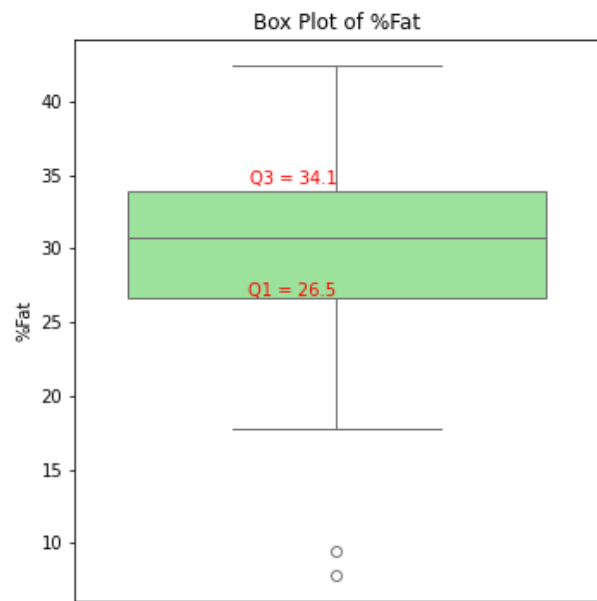
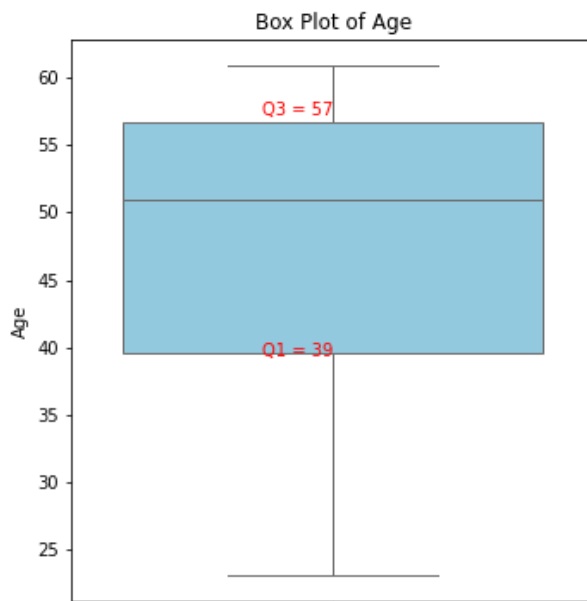
27.4	0.83	0.69
------	------	------

27.2	0.63	0.40
------	------	------

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
31.2	4.63	21.44
34.6	8.03	64.48
42.5	15.93	253.76
28.8	2.23	4.97
33.4	6.83	46.65
30.2	3.63	13.18
34.1	7.53	56.70
32.9	6.33	40.07
41.2	14.63	214.04
35.7	9.13	83.36

$$\text{Variance} = \frac{1544.12}{18} = 85.6$$

$$\text{Standard Deviation} = \sqrt{85.6} = 9.25$$



- 3) 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?

$$\text{Mean} = \frac{\sum \text{age values}}{\text{number of data points}} = \frac{809}{27} = 29.96$$

$$\text{Median} = 25$$

- b) What is the mode of the data?

The mode is the value that appears most frequently in the data

$$\text{Bimodal} = 25 \text{ and } 35$$

- c) What is the mid range of the data?

$$\text{Midrange} = \frac{13+70}{2} = 41.5$$

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

For 27 values:

$$Q1 \text{ (25th percentile)} = 20 \text{ (7th value)}$$

$$Q3 \text{ (75th percentile)} = 35 \text{ (21st value)}$$

- e) Give the five-number summary of the data.

$$\text{Minimum} = 13$$

$$Q1 = 20$$

$$\text{Median} = 25$$

$$Q3 = 35$$

$$\text{Maximum} = 70$$

f) Show a box plot of the data.

