BI Lab Assg No. 05

Atharva Yadav

T23-101

Q1] For the given data set

57, 57, 57, 58, 63, 66, 66, 67, 68 67, 68, 69, 70, 70, 70

72, 73, 75, 75, 76, 76, 78, 79, 81

(i) Mean:-

$$mean = \frac{\Sigma x_i}{N} = \frac{1590}{23} = 69.13$$

(ii) Median = x_{12} = 70

(iii) Mode = 57, 70

(iv) Midrange = $\frac{Min + Max}{2} = \frac{57 + 81}{2} = 69$

(v) Range = 81 - 57 = 24

(vi) Quartiles:-
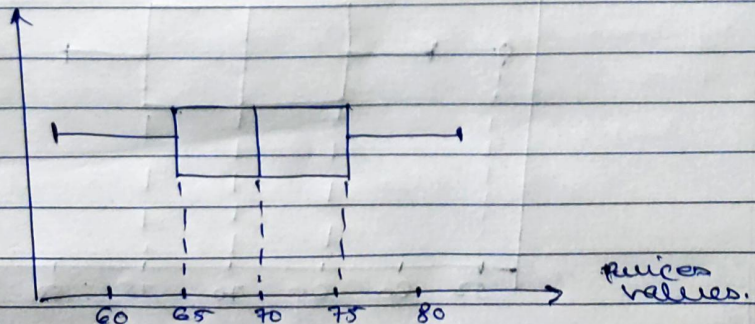
$Q_1 = 66$, $Q_2 = 70$, $Q_3 = 75$

IQR

(vii) IQR = $Q_3 - Q_1$ = 75 - 66 = 9

(viii) Box plot

60 65 70 75 80   → Prices
values.

Q2] 52, 57, 57, 58, 63, 66, 66, 67, 67, 68, 69, 70, 70, 70
72, 73, 75, 75, 76, 76, 78, 79, 89

(i) Mean $= \dfrac{\sum x_i}{N} = \dfrac{1593}{23} = 69.13 \quad 69.26$

(ii) Median $= X_{\frac{23+1}{2}} = X_{12} = 70$

(iii) Mode $= 70$

(iv) Midrange $= \dfrac{Min \times Max}{2} = \dfrac{52 + 89}{2} = 70.5$

(v) Range $= Max - Min = 89 - 52 = 37$

(vi) all quantities :-
$Q_1 = 66 \quad , Q_2 = 70, \quad Q_3 = 75$

(vii) IQR $= Q_3 - Q_1 = 9$

(viii) Box plot
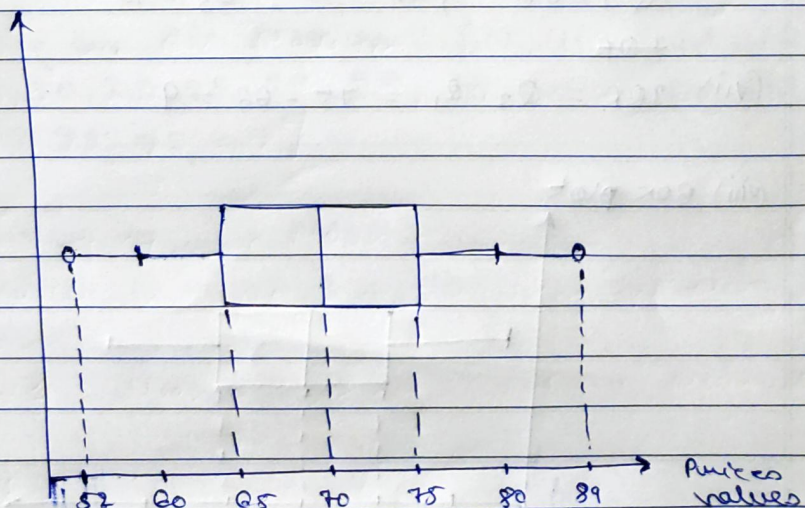
Q4]

Class interval	freq
0 - 10	4
10 - 20	6
20 - 30	18
30 - 40	34
40 - 50	40

(i) cumulative freq (CF):-

Class interval	freq	CF
0 - 10	4	4
10 - 20	6	10
20 - 30	18	28
30 - 40	34	62
40 - 50	40	102

(ii) Total freq $(N) = 102$

(iii) Median class $= N/2 = 102/2 = 51$

CF just exceeding 51 is 62 which correspond to interval 30-40

∴ Median class is 30-40

(iv) Median $= L + \dfrac{(N/2 - CF)}{F} \times h$

$= 30 + \left(\dfrac{102/2 - 28}{34} \right) \times 10$

$= 36.76$

L = lower bond of median class

h = class width.

Q5 Partition data into 4 bins using equi depth binning method and perform smoothing acc to following methods:

Data: 11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 49, 71, 72, 73, 75

Number of elements = 24

Bin size = $\frac{24}{6}$ = 0 data points per bin

~~Partition~~ Partitioning:-

Bin 1: [11, 13, 13, 15, 15, 16]
Bin 2: [19, 20, 20, 20, 21, 21]
Bin 3: [22, 23, 24, 30, 40, 45]
Bin 4: [45, 45, 71, 72, 73, 75]

a) Smoothing by Bin Median:-
Each value in a bin is replaced by the median of that bin.

Bin 1 : Median 15
Bin 2 : Median = 30
Bin 3 : Median = 30
Bin 4 :- Median = 72

Smoothing by Bin Medium: [15, 15, 15, 15, 15, 20, 20, 20, 20, 20, 20, 20, 30, 30, 30, 30, 30, 30, 72, 72, 72, 72, 72, 72]

b) Smoothing by Bin Mean:-
Each value in a bin is replaced by the mean of that bin.

Bin 1 : Mean = 13.83   Bin 4: Mean = 63.5
Bin 2 :- Mean = 90.17
Bin 3: Mean = 30.67

Smoothing by Bin Mean :- [13.83, 13.83, 13.83,
13.83, 20.17, 20.17, 20.17, 20.17, 30.67, 30.67,
30.67, 30.67, 63.5, 63.5, 63.5, 63.5]

c) Smoothing by Bin Boundaries:
Each value in a bin is replaced by the closest
boundary value (max or mini of the bin)

Bin 1:- Boundary values = 11 (min), 16 (max)
(11; 13 → 11, 13 → 11, 15 → 16, 15 → 16,]
smoothed Bin : { 11, 11, 11, 16, 16, 16 }

Bin 2:- Boundary values = 19 (min), 21 (max)
(19, 20 → 19, 20 → 19, 20 → 19, 21, 21]

Bin 3:- Boundary values = 22 (min), 45 (max)
(22, 23 → 22, 24 → 22, 30 → 45, 40 → 45, 45]
smoothed Bin 3: { 22, 22, 22, 45, 45, 45 }

smoothly by Bin Boundaries:-
(11, 11, 11, 16, 16, 16, 19, 19, 19, 19, 21, 21,
22, 22, 22, 45, 45, 45, 45, 45, 75, 75, 75, 75]

3. Use MS-Excel to

a. Extract summary of descriptive statistics

Mean	29.96296
Median	25
Mode	25
Range	57
Standard Deviation	12.94212
Variance	167.4986

b. percentiles of 10, 20,25, 50,75, 80,90,100

10th Percentile	16
20th Percentile	20
25th Percentile	20.5
50th Percentile	25
75th Percentile	35
80th Percentile	35.8
90th Percentile	45.4
100th Percentile	70

c. Create frequency bin
d. Draw histogram of the following dataset

Bin	frequency
20	7
30	8
40	8
50	2
60	1
70	1

Histogram of Given Dataset
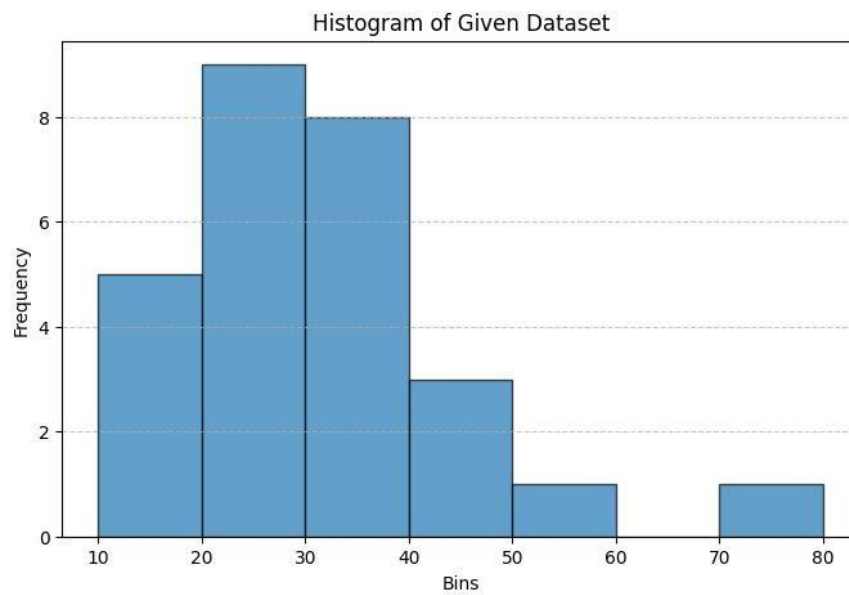
e. Draw boxplot of the following dataset

Boxplot of Given Dataset