



# COMSATS University Islamabad, Wah Campus

## Department of Computer Science

Course : Introduction to Data Science  
Class : SE-7  
Semester: FALL 2023

Total Marks : 10  
Dated : 20/09/2023  
Submission Deadline: (24-Sept-2023) 11:59 PM.

**Name: Muhammad Ali Ahmad**

**Reg no.: FA20-BSE-006**

**Submitted To: Ma'am Maha Rasheed**  
**Assignment (01) (CLO 1)**

### Instructions:

- Make sure to read all instructions
- Make sure this page is the first page of your assignment file.
- All answers should start from page 2.
- Assignment should be handwritten, scanned and submitted in **PDF** format.
- Mention both name and roll number on this page.
- Give justifications where necessary!
- MARKS WILL BE DEDUCTED ON NOT FOLLOWING THE FORMAT.

### Assignment Tasks:

You are tasked with analyzing a raw dataset containing the ages of a group of individuals. Your goal is to apply bin smoothing techniques, specifically 'Smoothing by Mean' and 'Smoothing by Boundary,' to improve the representation of the data. Preprocess the data into bins of frequency 10 as necessary.

**Dataset: Ages of Individuals**

40, 23, 31, 27, 19, 45, 36, 20, 29, 22, 43, 38, 25, 18, 47, 27, 35, 30, 42, 33, 26, 21, 43, 28, 32, 44, 19, 41, 37, 18, 49, 34, 23, 22, 46, 18, 20, 50, 39, 27, 19, 31, 27, 22, 45, 38, 23, 18, 47, 36.

DATE: 23/09/2023

## ASSIGNMENT

JIDS

NAME:- Muhammad Ali Ahmad

Reg no:- FH20-BSE-006

You are tasked .....

Sort Dataset into Ascending Order

18, 18, 18, 18, 19, 19, 19, 20, 20, 21,  
22, 22, 22, 23, 23, 23, 25, 26, 27, 27,  
27, 27, 28, 29, 30, 31, 31, 32, 33, 34,  
35, 36, 36, 37, 38, 38, 39, 40, 41, 42,  
43, 43, 44, 45, 45, 46, 47, 47, 49, 50

Equi-Depth bins:-

- Bin 1: 18, 18, 18, 18, 19, 19, 19, 20, 20, 21.
- Bin 2: 22, 22, 22, 23, 23, 23, 25, 26, 27,  
27.
- Bin 3: 27, 27, 28, 29, 30, 31, 31, 32, 33, 34
- Bin 4: 35, 36, 36, 37, 38, 38, 39, 40, 41, 42
- Bin 5: 43, 43, 44, 45, 45, 46, 47, 47, 49,  
50

Smoothing by bin mean:-

Bin 1 (mean):-

$$\text{mean} = \frac{18+18+18+18+19+19+19+20+20+21}{10}$$

$$= \frac{190}{10}$$

$$\text{mean} = 19$$

Bin 2 (mean):-

$$\text{mean} = \frac{22+22+22+23+23+23+25+26+27+27}{10}$$

$$= \frac{240}{10} \Rightarrow 24$$

DATE: \_\_\_/\_\_\_/\_\_\_

Bin 3:- (mean)

$$\text{mean} = \frac{27 + 27 + 28 + 29 + 30 + 31 + 31 + 32 + 33 + 34}{10}$$

$$= \frac{302}{10}$$

$$\text{mean} = 30.2 \approx 30$$

Bin 4 (mean):-

$$\text{mean} = \frac{35 + 36 + 36 + 37 + 38 + 38 + 39 + 40 + 41 + 42}{10}$$

$$= \frac{382}{10}$$

$$\text{mean} = 38.2 \approx 38$$

Bin 5 (mean):-

$$\text{mean} = \frac{43 + 43 + 44 + 45 + 45 + 46 + 47 + 47 + 49 + 50}{10}$$

$$= \frac{459}{10}$$

$$\text{mean} = 45.9 \approx 46$$

So the smoothing by bin mean are

- Bin 1: 19, 19, 19, 19, 19, 19, 19, 19, 19, 19
- Bin 2: 24, 24, 24, 24, 24, 24, 24, 24, 24, 24
- Bin 3: 30, 30, 30, 30, 30, 30, 30, 30, 30, 30
- Bin 4: 38, 38, 38, 38, 38, 38, 38, 38, 38, 38
- Bin 5: 46, 46, 46, 46, 46, 46, 46, 46, 46, 46



DATE: \_\_\_/\_\_\_/\_\_\_

### Smoothing by bin Boundary:-

- Bin 1: 18, 18, 18, 18, 18, 18, 18, 21,  
21, 21
  - Bin 2: 22, 22, 22, 22, 22, 22, 22, 27,  
27, 27
  - Bin 3: 27, 27, 27, 27, 27, 34, 34, 34,  
34, 34
  - Bin 4: 35, 35, 35, 35, 35, 35, 42, 42,  
42, 42
  - Bin 5: 43, 43, 43, 43, 43, 43, 50, 50,  
50, 50
-