# STATISTICS

# Lecture II

# Sturge's Rule

Sturge's Rule is usually used to determine the number of classes. In symbols

$$k = 1 + 3.3 \log_{10} N$$

where k = number of classes

N = number of observations

$$Common\ class\ width = \frac{class\ range}{number\ of\ class}$$

where

class range = Maximum value – minimum value

## Example

The following data are the final examination marks in statistics of 40 students at the institute of Economics.

**Raw Data** **Collect**

68 , 79 , 67 , 63 , 93 , 74 , 78 , 87 , 73 , 86 ,

62 , 68 , 75 , 77 , 79 , 61 ,61 , 75 , 71 , 71 ,

72 , 88 , 65 , 80 , 66 , 82 , 60 , 53 , 60 , 62 ,

76 , 67 , 94 , 89 , 76 , 85 , 71 , 73 , 94 , 86

| f | Stem | Leaf |
|---|---|---|
| 1 | 5 | 3 |
| 13 | 6 | 8 7 3 2 8 1 1 5 6 0 0 2 7 |
| 15 | 7 | 9 4 8 3 5 7 9 5 1 1 2 6 6 1 3 |
| 8 | 8 | 7 6 8 0 2 9 5 6 |
| 3 | 9 | 3 4 4 |

| f | Stem | Leaf |
|---|---|---|
| 1 | 5 | 3 |
| 13 | 6 | 0 0 1 1 2 3 5 6 7 7 8 8 |
| 15 | 7 | 1 1 1 2 3 3 4 5 5 6 6 7 8 9 |
| 8 | 8 | 0 2 5 6 6 7 8 9 |
| 3 | 9 | 3 4 4 |

# Organization

53 , 60 , 60 , 61 , 61 , 62 , 62 , 63 , 65 , 66 ,

67 , 67 , 68 , 68 , 71 , 71 , 71 , 72 , 73 , 73 ,

74 , 75 , 75 , 76 , 76 , 77 , 78 , 79 , 79 , 80 ,

82 , 85 , 86 , 86 , 87 , 88 , 89 , 93 , 94 , 94

class range = 94 – 53 = 41

## Sturge's Rule

$$k = 1 + 3.3 \log_{10} N$$

# Sturge's Rule

$$k = 1 + 3.3 \log_{10} N$$

$$k = 1 + 3.3 \log_{10} 40$$

$$k = 1 + 3.3 \times 1.6021$$

$$k = 6.2869 \approx 7$$

$$k = 7$$

$$Common\ class\ width = \frac{41}{7} = 5.85 \approx 6$$

| Marks | Tally | Frequency |
|---|---|---|
| 53 – 58 | | 1 |
| 59 – 64 | | 7 |
| 65 – 70 | | 6 |
| 71 – 76 | | 11 |
| 77 – 82 | | 6 |
| 83 – 88 | | 5 |
| 87 – 94 | | 4 |
| Total | | 40 |

| Marks | $f_j$ | $X_j$ | $u_j$ | $f_j u_j$ | $F_j$ | $u_j^2$ | $f_j u_j^2$ |
|---|---|---|---|---|---|---|---|
| 53 – 58 | 1 | 55.5 | – 3 | – 3 | 1 | 9 | 9 |
| 59 – 64 | 7 | 61.5 | – 2 | – 14 | 8 | 4 | 28 |
| 65 – 70 | 6 | 67.5 | – 1 | – 6 | 14 | 1 | 6 |
| 71 – 76 | 11 | 73.5 | 0 | 0 | 25 | 0 | 0 |
| 77 – 82 | 6 | 79.5 | 1 | 6 | 31 | 1 | 6 |
| 83 – 88 | 5 | 85.5 | 2 | 10 | 36 | 4 | 20 |
| 89 – 94 | 4 | 91.5 | 3 | 12 | 40 | 9 | 36 |
| | N = 40 | | $\sum fu =$ | 5 | | $\sum fu^2 =$ | 105 |

$$c = 6 \ , \quad A = 73.5$$

# Measure of central Tendency

## Mean

$$\overline{X} = \frac{\sum f x}{N}$$

$$\overline{X} = A + \frac{\sum f d}{N}$$

$$, d_j = X_j - A \ , \ A = Assume \ \ mean$$

## For Group Data

$$\overline{X} = A + \frac{\sum f u}{N} \times c$$

$$, u_j = \frac{d_j}{c} = \frac{X_j - A}{c} \ , \ c = class \ \ width$$

## Mean

$$\overline{X} = A + \frac{\sum fu}{N} \times c$$

$$\overline{X} = 73.5 + \frac{5}{40} \times 6 = 73.5 + 0.75$$

$$\overline{X} = 74.25$$

# Median

$$\overline{X} = L_1 + \frac{\frac{N}{2} - \left(\sum f\right)_1}{f_{median}} \times c$$

$L_1$ = Lower boundary of median class

$f_{median}$ = frequency of median class

$(\sum f)_1$ = the sum of frequencies of lower than the median class (or) cumulative frequency of preceding the median class

c = class width

# Median

$$Median = L_1 + \frac{\frac{N}{2} - \left(\sum f\right)_1}{f_{median}} \times c$$

$$Median = 70.5 + \frac{20 - 14}{11} \times 6$$

$$Median = 70.5 + 3.27 = 73.77$$

# Quartiles $Q_1, Q_2, Q_3$

$$Q_1 = 25\% \text{ of set, value of } \frac{N}{4} \text{ time}$$

$$Q_2 = 50\% \text{ of set} = \text{median, value of } \frac{2N}{4} \text{ time}$$

$$Q_3 = 75\% \text{ of set, value of } \frac{3N}{4} \text{ time}$$

$$Q_1 = L_{Q_1} + \frac{\frac{N}{4} - \left(\sum f\right)_{Q_1}}{f_{Q_1}} \times c$$

$$\text{Inter quartile range} = Q_3 - Q_1$$

$$\text{Semi-inter quartile range (quartile deviation)} = \frac{Q_3 - Q_1}{2}$$

$$\frac{3N}{4} = \frac{3 \times 40}{4} = 30$$

$$Q_3 = L_{Q_3} + \frac{\frac{3n}{4} - \left(\sum f\right)_{Q_3}}{f_{Q_3}} \times c$$

$$Q_3 = 76.5 + \frac{30 - 25}{6} \times 6 = 81$$

## Deciles

$$D_1, D_2, \ldots\ldots, D_9$$

$$D_1 = 10\% \ of \ set, Value \ of \ \frac{N}{10} \ time$$

–

–

–

$$D_9 = 90\% \ of \ set, \ value \ of \ \frac{9N}{10} \ time$$

$$D_8 = L_{D_1} + \frac{\frac{8N}{10} - \left(\sum f\right)_{D_8}}{f_{D_8}} \times c$$

## Percentiles

$$P_1 , P_2 , \ldots\ldots , P_{99}$$

$$P_1 = 1\% \ of \ set, Value \ of \ \frac{N}{100} \ time$$

–

–

–

$$D_{99} = 99\% \ of \ set, \ value \ of \ \frac{99N}{100} \ time$$

$$P_{28} = L_{P_{28}} + \frac{\frac{28N}{100} - \left(\sum f\right)_{P_{28}}}{f_{P_{28}}} \times c$$

$$semi \ 10 - 90 \ percentile \ range = \frac{P_{90} - P_{10}}{2}$$

# Mode

$$Mode = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

**$L_1$ = Lower boundary of modal class**

$\Delta_1$ = *the different between the frequency of the madal class and the preceding class*

$\Delta_2$ = *the different between the frequency of the madal class and the suceeding class*

$$L_1 = 70.5$$

$$\Delta_1 = 11 - 6 = 5 \qquad \Delta_2 = 11 - 6 = 5$$

$$Mode = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

$$Mode = 70.5 + \frac{5}{5+5} \times 6 = 73.5$$

# Variance

$$Variance = \frac{\sum f(X_j - \overline{X})^2}{N}$$

$$Variance = \left( \frac{\sum f d^2}{N} - \left( \frac{\sum f d}{N} \right)^2 \right)$$

# For Group Data

$$Variance = \left( \frac{\sum f u^2}{N} - \left( \frac{\sum f u}{N} \right)^2 \right) \times c^2$$

# Variance

$$Variance = \left( \frac{\sum fu^2}{N} - \left( \frac{\sum fu}{N} \right)^2 \right) \times c^2$$

$$Variance = \left( \frac{105}{40} - \left( \frac{5}{40} \right)^2 \right) \times 36$$

# Standard Deviation

$$s = \sqrt{\frac{\sum f(X_j - \overline{X})^2}{N}} \quad , \quad s = \sqrt{Variance}$$

$$s = \sqrt{\left( \frac{f d^2}{N} - \left( \frac{f d}{N} \right)^2 \right)}$$

# For Group Data

$$s = \sqrt{\left( \frac{f u^2}{N} - \left( \frac{f u}{N} \right)^2 \right)} \times c$$

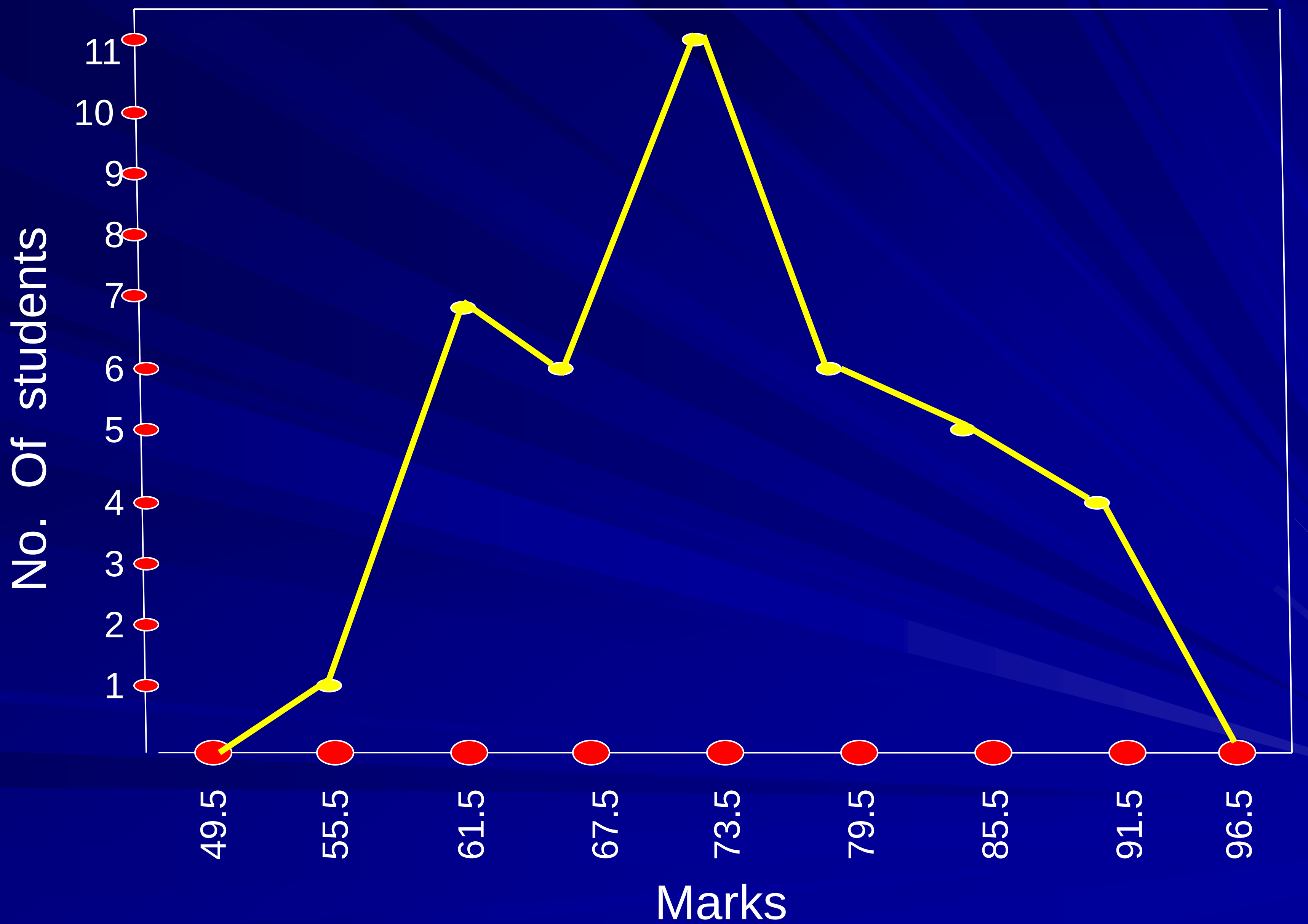
# Standard Deviation

$$s = \sqrt{\left( \frac{f u^2}{N} - \left( \frac{f u}{N} \right)^2 \right)} \times c$$

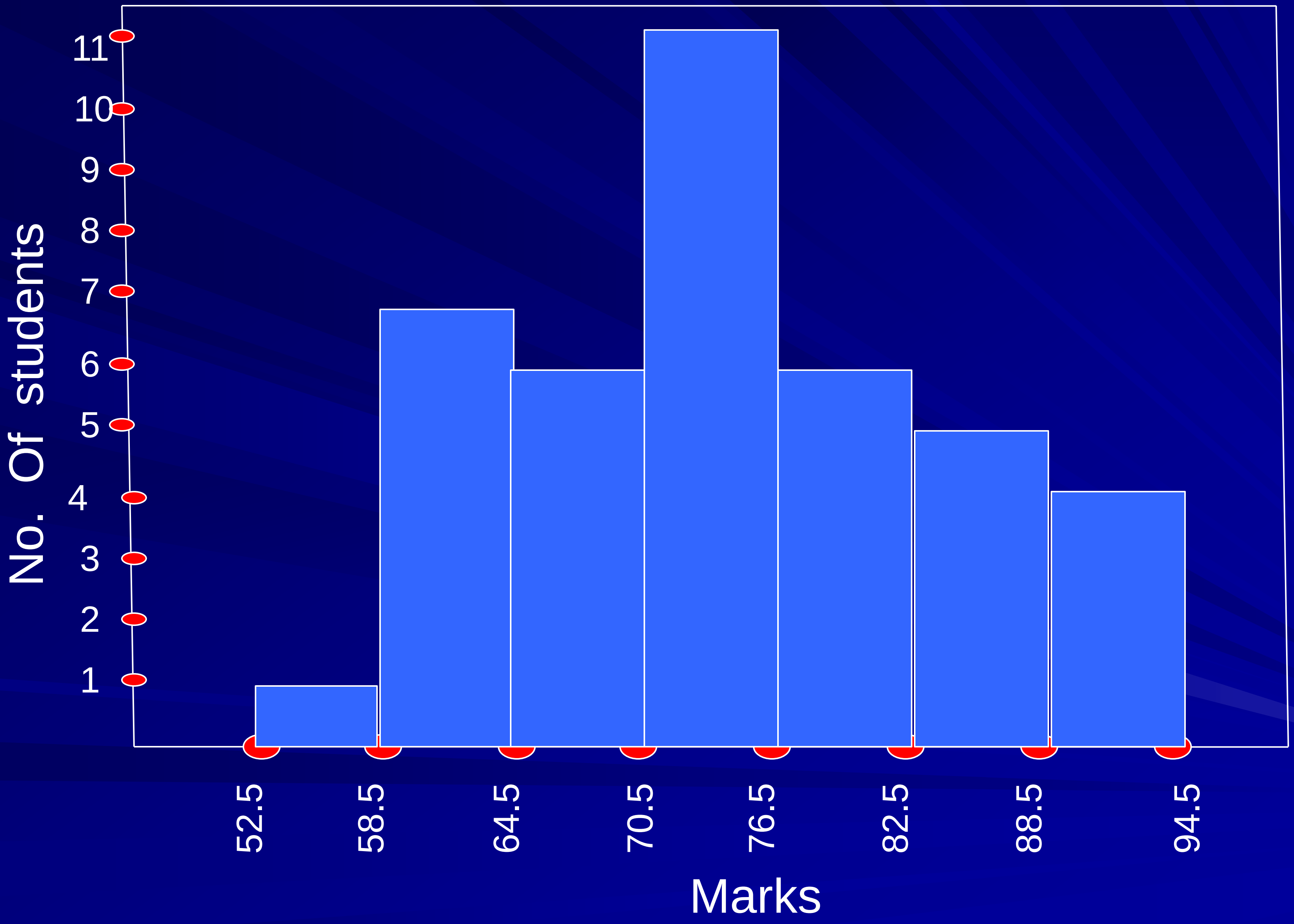
$$s = \sqrt{\left( \frac{105}{40} - \left( \frac{5}{40} \right)^2 \right)} \times 6$$

| Marks | $f_j$ | $F_j$ |
| --- | --- | --- |
| 53 – 58 | 1 | 1 |
| 59 – 64 | 7 | 8 |
| 65 – 70 | 6 | 14 |
| 71 – 76 | 11 | 25 |
| 77 – 82 | 6 | 31 |
| 83 – 88 | 5 | 36 |
| 87 – 94 | 4 | 40 |
| Total | 40 | |

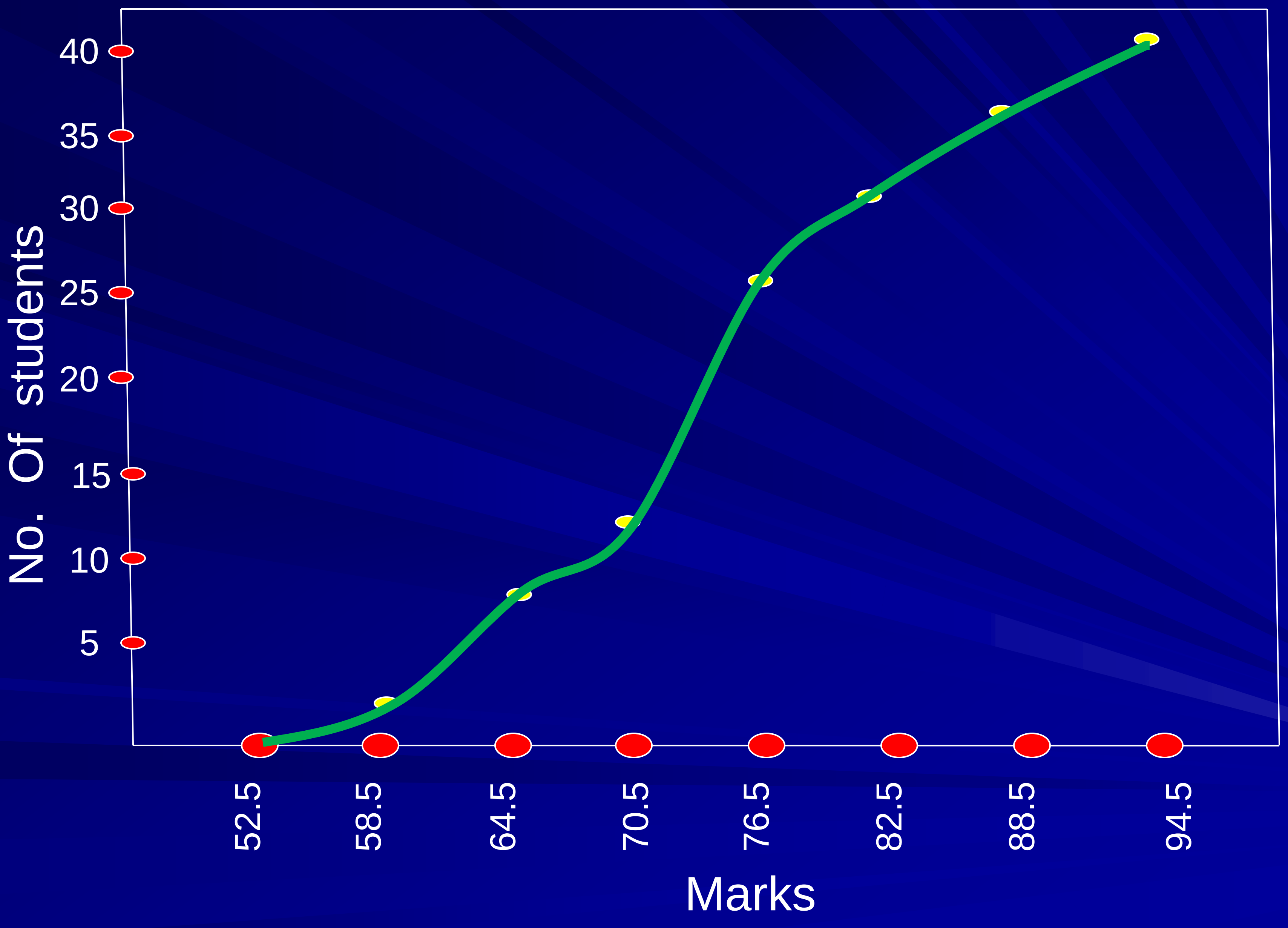
# Frequency Polygon



# Frequency Polygon



# Cumulative Frequency Ogive



# Conclusion