Classification or Organization

Classification is the process of arranging data into sequences and groups according to their common characteristics

In general the data can be classified on the following four bases.

- 1.Geographical
- 2.Chronological
- 3. Qualitative
- 4. Quantitative

Here we will study only Quantitative data

Variables : There are two variables in Statistics

Discrete	and	Continuous
The variable which has called Discrete variabl	s no fraction or decima l e . Ex : No. of death in	al value is a city ,No. of bi

rth

in government hospital, Size of shoe, size of shirtetc.

The variable which has fraction or decimal part is called **Continuous variable**. Ex : Height, Weightetc.

Frequency Distribution (Uni-variate Frequency Distribution)

The organization of the data on the basis of quantitative classification involves the following four stages :

I. The set of individual observations.

II. Discrete or ungrouped frequency distribution.

III.Grouped frequency distribution.

IV.Continuous frequency distribution.

i. The set of individual observations

Let us see one example

Following marks obtained by 50 students during a test in a class (Max Mark 10)

6	5	8	4	5	3	6	2	1	7
3	0	5	4	8	2	9	7	1	2
4	3	6	1	7	6	5	2	8	6
9	2	0	3	6	8	7	4	1	2
5	6	5	4	6	3	2	7	5	6

The above data is in the form of individual observation.

Now how to change this data in tabular form ? It means how to organize this data ?

ii. Discrete or ungrouped frequency distribution

Marks	Tallies	No. of Students(f)
0		2
1		4
2	1111	7
3	1	5
4	1	5
5	1	7
6	₩U IIII	9
7	++11	5
8		4
9	11	2
		Total=50

We changed above individual observations in tabular form ,where 0 comes 2 times ,1 comes 4 times and so on ...

The individual observations are 50 so the total frequency is 50.

Basic terms Related to Grouped and Continuous Frequency Distribution.

(1) No of classes: 8

Daily wages(in Rs.)	No. of Workers(f)
75-80	9
80-85	12
85-90	15
90-95	11
95-100	20
100-105	20
105-110	11
110-115	2

(2)Class-limits: Two class limits

(a) Lower limit 75(b)Upper limit 80

(2) Class-width or

Magnitude of Class-interval(i):

80-75=5 (3) Mid-point or Class-mark : (75+80)/2=77.5

iii. Grouped Frequency Distribution

Ex: Marks obtained by 100 students

15	21	29	30	50	34	36	38	46	41
15	21	29	30	33	34	37	38	46	41
16	22	29	30	33	34	37	38	40	42
16	22	28	31	33	35	37	39	40	42
17	23	28	31	33	35	47	39	40	43
17	23	27	31	51	35	48	39	40	43
17	24	27	32	34	35	49	39	40	44
18	24	27	32	34	35	49	53	54	44
19	25	26	32	52	36	38	54	41	44
20	25	26	50	52 By Ashish	36 Mishra	38	54	41	45

Now let us apply the Discrete or ungrouped frequency distribution. Here minimum value is 15 and maximum value 54.

Mark s	Tally Bars	Frequ e-ncy	Mark s	Tally Bars	Frequ e-ncy	Marks	Tally Bars	Frequ e-ncy	Mark s	Tally Bars	Frequ e-ncy
15	П	2	25	Ш	2	35	1111	5	45	1	1
16	П	2	26	П	2	36	Ш	3	46	П	2
17	Ш	3	27	Ш	3	37	Ш	3	47	I.	1
18	I	1	28	П	2	38	1111	5	48	I.	1
19	1	1	29	Ш	3	39	Ш	4	49	П	2
20	1	1	30	Ш	3	40	1111	5	50	П	2
21	П	2	31	Ш	3	41	Ш	4	51	I.	1
22	П	2	32	Ш	3	42	П	2	52	П	2
23	П	2	33		4	43	П	2	53	I.	1
24	П	2	34	1111	5	44	Ш	3	54		3

In the above frequency distribution no of rows are 40which is too high.

So here we will apply grouped frequency distribution.

Marks	Tally Bars	No of Students(f)
15-19	+++11111	9
20-24	1+++1111	9
25-29	7+++17+++11	12
30-34	1+++1+++++111	18
35-39	++11++11++11++11	20
40-44	-+++1++++++1	16
45-49	1111	7
50-54	1++1	9
		Total=100

The above frequency distribution has only 8 rows which looks good.So the discrete data set which has high range(Max value - Min value) we will apply grouped frequency distribution.
Note: The above frequency distribution also known as inclusive series or inclusive frequency distribution.

(iv)Continuous Frequency Distribution Ex: daily wages of 100 workers in a factory (in Rs.)

75	80	84	94	96	88	86	92	87	109
75	80	100	95	87	104	93	104	90	86
76	81	110	101	99	103	89	103	114	94
76	82	93	99	107	100	100	101	104	107
77	82	109	89	99	98	101	103	101	86
78	83	95	102	97	87	102	100	107	107
78	83	100	97	98	100	99	88	97	88
79	84	89	93	93	96	86	106	91	93
79	84	106	97	106	95	105	98	103	85
80	84	96	102	94	86	97	108	98	98

In the above data min value is 75 and max value is 114.

Frequency Distribution of Wages of 100

Workers in a Factory

Daily wages(in Rs.)	Tally Marks	No. of Workers(f)
75-80	1+++1	9
80-85	1+++1 1++1 11	12
85-90	<u>+++1 ++1 ++1</u>	15
90-95	144114411	11
95-100	+++++++++++++++++++++++++++++++++++++++	20
100-105	1+++ 1+++ 1+++	20
105-110	1111111	11
110-115	11	2
		Total=100

Note: The above frequency distribution also known as exclusive series or exclusive frequency distribution.

Calculating No of Classes & Class-width

No. of Classes :

Two basic questions.

- (a) If no. of classes are too small ?...it will not look good
- (b) If no. of classes are too large?....it will not look good

So in general we want to put 8 to 10 classes.

We can also calculate No of classes by below rule

according to Sturges rule $k= 1+ 3.322 \log_{10} N$ Where k=no of classes N= total frequency If N=100 then $k = 1 + 3.322 \log_{10} 100$ $= 1 + 3.322 \log_{10} 10^{2}$ $= 1 + 3.322 \times 2 \log_{10} 10$ $= 1 + 3.322 \times 2$ = 1 + 6.644 $= 7.644 \approx 8$ No. of Classes = 8

Size of class Interval or class width(i) :

"Size of class interval is inversely proportional to the no. of classes"

i = Range = Range No. of classes $1 + 3.322 \log_{10} N$ Range = $X_{max} - X_{min}$ If k = 8 and Range = 114-75=39then i = $39/8 = 4.875 \approx 5$

Open end classes

Class-Interval	frequency
Below 20	16
20-40	20
40-60	30
60-80	26
80-100	14
100 and above	12

<u>Conversion of Inclusive series into Exclusive</u> <u>series</u>

First of all take an Inclusive series

Marks	No of Students(f)
15-19	9
20-24	9
25-29	12
30-34	18
35-39	20
40-44	16
45-49	7
50-54	9
	Total=100

As we know in inclusive series the difference between upper limit of class and lower limit of succeeding class is 1 so

Step 1:

Calculate the difference between upper limit of a class and lower limit of its succeeding class

Step 2:

Now divide this value by 2

$$1/2 = 0.5$$

Step 3:

Now subtract the value find in step 2 by all the lower limits and add this value in all the upper limit .

So we will get the final table as follows.

Marks	No of Students(f)
14.5-19.5	9
19.5-24.5	9
24.5-29.5	12
29.5-34.5	18
34.5-39.5	20
39.5-44.5	16
44.5-49.5	7
49.5-54.5	9
	Total=100

Write the frequency as it is.