Chapter (2)

Frequency Distribution

A frequency distribution lists each category of data and the number of occurrences for each category of data.

Frequency Distribution for Qualitative Data

Example (1): A sample 30 persons who often consume donuts were asked what variety of donuts is their favorite. The responses from these 30 persons are as of follows:

glazed	filled	other	plain	glazed	other
frosted	filled	filled	glazed	other	frosed
glazed	plain	other	glazed	glazed	filled
frosed	plain	other	other	frosed	filled
filled	other	frosted	glazed	glazed	filled

Construct a frequency distribution table for these data.

Solution:

Table (1) Frequency Distribution of Favorite Donut Variety.

Donut Variety	Tally	Frequency (f)
Glazed	𝍸 |||	8
Filled	𝍸 ||	7
Frosted	𝍸	5
Plain	|||	3
Other	𝍸 ||	7
		Sum = 30

Numeric frequency : $f \rightarrow \sum f_i = N$

Relative frequency : $f_r = \frac{f}{N} \rightarrow \sum f_r = 1$

Percent frequency : $f_p = f_r * 100 \rightarrow \sum f_p = 100$

Example (2): Determine the relative frequency and percentage distribution for data in table (1).

Solution:

Table (2): Relative Frequency and Percentage Distributions of Favorite Donut Variety

Donut Variety	Relative Frequency	Percentage
Glazed	8/30 = .267	.267(100) = 26.7
Filled	7/30 = .233	.233(100) = 23.3
Frosted	5/30 = .167	.167(100) = 16.7
Plain	3/30 = .100	.100(100) = 10.0
Other	7/30 = .233	.233(100) = 23.3
	Sum = 1.000	Sum = 100%

Frequency Distribution for Quantitative Data

A frequency distribution for quantitative data lists all the classes and the numbers of values that belong to each class, Data presented in the form of frequency distribution are called **grouped data**. Whereas the data that list undivided value are called **ungrouped data**.

- Frequency distribution into classes:

The sample range is sub-divided in to a number of classes. Usually:

For size
- > 50 : 10-20 classes
- ≤ 50 : 5-10 classes are used

- Sample range = Largest - smallest

- class limits = highest and lowest measurements in the class.

- class interval = upper limit – lower limit.

- class boundaries = limits $\mp \frac{1}{2}$ unit in LSD.

- class width = upper – lower boundaries.

- class mark = mid point of class.

- e. g. if the class limits are 670 → 678.

* class interval = 678 – 670 = 8.

* class boundaries are : lower bound. = 670 – 0.5 = 669.5

upper bound. = 678 + 0.5 = 678.5

* Class width = 678.5 – 669.5 = 9

* class mark = $\frac{670+678}{2}$ = 674

Or :

Class mark = $\frac{669.5+678.5}{2}$ = 674

e. g. if class limits are 5.87 → 6.32 :

* class interval = 6.32 – 5.87 = 0.45

* class boundaries are : lower bound. = 5.87 – 0.005 = 5.865.

upper bound. = 6.32 + 0.005 = 6.325.

* Class width = 6.325 – 5.865 = 0.46

* class mark = $\frac{5.87+6.32}{2}$ = 6.095

Or :

Class mark = $\frac{5.865+6.325}{2}$ = 6.095

Determination of classes :

1. Determine the range.
2. Determine the total width

   Total width = range + one unit in LSD.
3. Divided the total width into a convenient no. of classes

   Class width = $\frac{total\ width}{no.of\ classes}$

Note :

(Adjust the total width by adding one or two units in LSD if necessary, to select a suitable no. of classes, so that the class width is of a similar accuracy to the measurements).

4. Determine class interval :

   Class interval = class width – one unit in LSD
5. Starting at lowest measurements, calculate the limits of successive classes.

Example :

The life of electric bulbs in hours was sampled :

690	701	722	684	680
728	705	693	691	688
740	663	676	738	714
698	687	703	726	699
694	705	717	682	717
712	733	705	673	694
679	680	664	691	669
689	702	710	696	697
685	724	726	698	688
702	696	708	696	710

Solution

- sample size = 50 measurements.

1. range = 740 – 663 = 77 hr

2. one unit in LSD = 1

   ∴ total width = 77 +1 = 78

3. select no. of classes (for example take 5 classes) so the class width $= \frac{78}{5} = 15.6$

Which is not the same accuracy as the data.

So take 6 classes :

class width $= \frac{78}{6} = 13$

Which is the same accuracy as the data.

4. class interval = 13 – 1 = 12

Freq. dist. Table

	Class limit	Class bound.	Class mark	Freq.
1.	663-675	662.5-675.5	669	4
2.	676-688	675.5-688.5	682	10
3.	689-701	688.5-701.5	695	15
4.	702-714	701.5-714.5	708	11
5.	715-727	714.5-727.5	721	6
6.	728-740	727.5-740.5	734	4
				N = 50

Types of Frequency :

1. Numeric frequency : $f \rightarrow \sum f_i = N$

2. Relative frequency : $f_r = \frac{f}{N} \rightarrow \sum f_r = 1$

3. Percent frequency : $f_p = f_r * 100 \rightarrow \sum f_p = 100$

4. Cumulative frequency : The freq. is also expressed cumulatively of : f , f_r , f_p.

- Cumulative freq. of class K is the sum of frequencies of all classes up to K.

$$fc_K = \sum_{i=1}^{K} f_i \quad , \quad fc_{rK} = \sum_{i=1}^{K} f_{ri} = 1$$

$$fc_{pK} = \sum_{i=1}^{K} f_{pi} = 100$$

f	f_r	f_p
4	0.08	8
10	0.2	20
15	0.3	30
11	0.22	22
6	0.12	12
4	0.08	8
$\sum f = 50$	$\sum f_r = 1$	$\sum f_p = 100$

$$* fc = \sum_{i=1}^{i=6} f_i = 50$$

$$* fc_r = \sum_{i=1}^{6} f_{ri} = 1$$

$$* fc_p = \sum_{i=1}^{6} f_{pi} = 100$$

Graphical presentation of freq. dist. :

- Classified data may be presented as graphical plot with freq. as vertical axis Versus measurement as horizontal axis :

1) Histogram : Is a bar chart, in which each class is represent by a rectangle, whose base extends between the class boundaries and the area proportional to frequency.

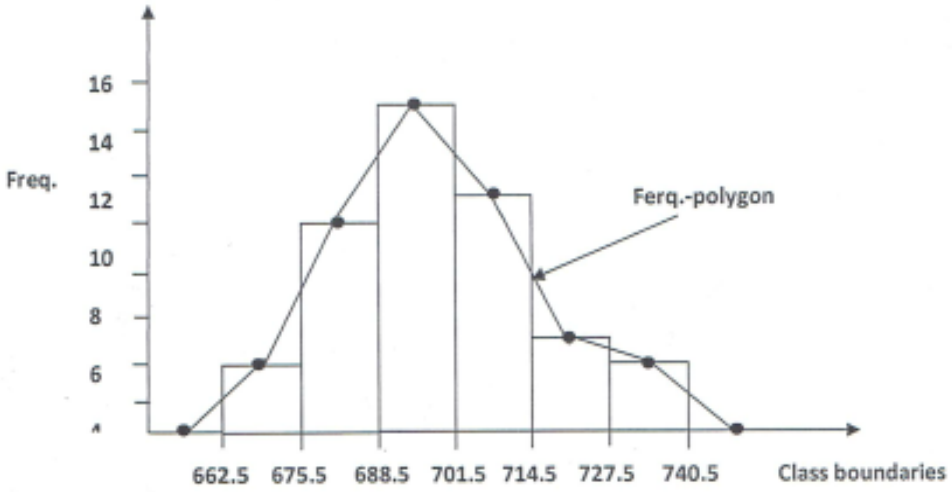
2) Frequency polygon : Consists of lines joining the class mark with freq., it may be obtained from the histogram by joining the mid-points of the bar tops.

3) Frequency curve : Is a smoothed frequency polygon in to a continuous curve.

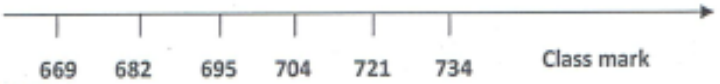
4) Cumulative freq. curve (Ogive) : Is a smoothed cumulative

Graphical presentation of frequency dist. Table :

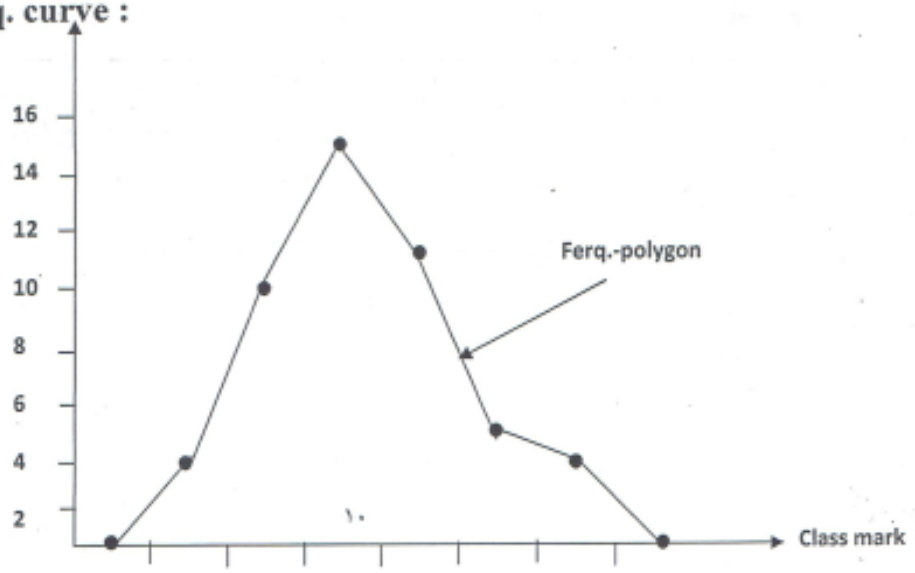
1) Histogram :



2) freq. polygon :

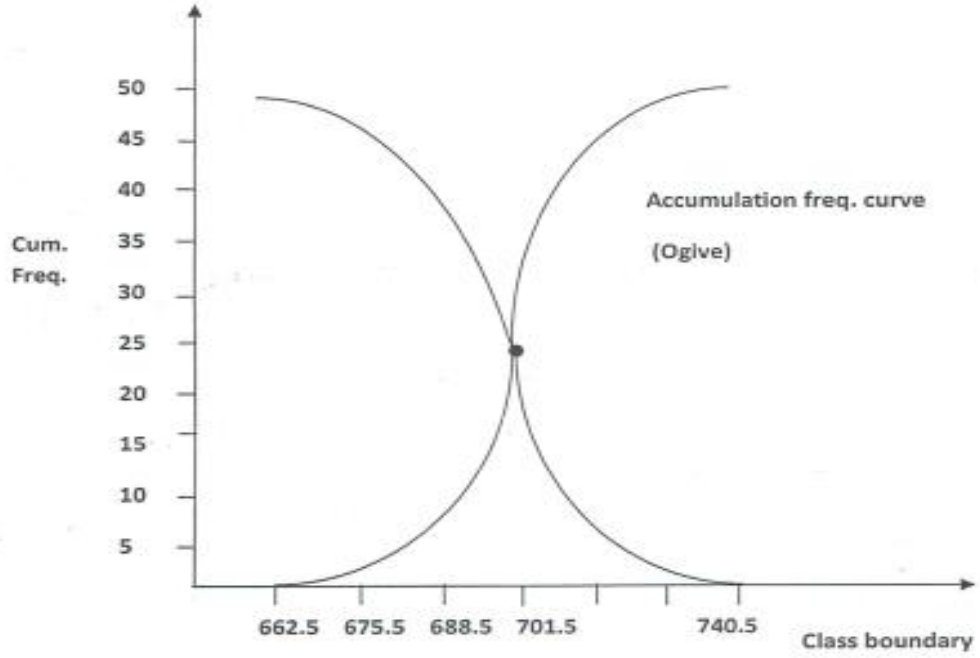


3) freq. curve :



4) Cumulative freq. :

Ascending cum. Freq.		Desending cum.	
Upper boundaries	cum. Freq.	Upper boundaries	cum. Freq.
Less than 662.5	0	Greater than 662.5	50
Less than 675.5	4	Greater than 675.5	46
Less than 688.5	14	Greater than 688.5	36
Less than 701.5	29	Greater than 701.5	21
Less than 714.5	40	Greater than 714.5	10
Less than 727.5	46	Greater than 727.5	4
Less than 740.5	50	Greater than 740.5	0



Tutorial Sheet No. (1)

For the following data groups obtain :

1) Frequency distribution table.

2) f_r , f_p , f_c

3) Histogram, freq. polygon, and Ogives

Data 1)

6.3	7.0	7.5	9.0	7.7	7.8	7.1	8.1
6.6	7.2	8.3	8.5	6.9	7.7	8.0	7.3
8.6	7.1	8.7	6.4	7.7	7.4	8.0	7.6
7.5	7.2	7.5	8.8	7.8	7.9	7.3	7.0
6.8	8.1	8.4	6.7	7.1	8.2	8.1	7.7

Data 2)

5.4	4.1	5.2	2.8	4.9	5.6	4.0	4.1	4.3
3.9	4.5	6.1	3.7	2.3	4.5	4.9	5.6	4.3
4.2	3.2	5.0	4.8	3.7	4.6	5.5	1.8	5.1
5.1	6.3	3.3	5.8	4.4	4.8	3.0	4.3	4.7

Data 3)

12.16	12.38	12.21	12.55	12.22	12.40	12.43	12.35
12.31	12.07	12.31	12.33	12.56	12.41	12.42	12.44
12.30	12.39	12.10	12.37	12.18	12.48	12.19	12.43
12.25	12.37	12.47	12.49	12.35	12.28	12.30	12.31
12.35	12.20	12.39	12.54	12.59	12.29	12.46	12.09