Frequency Distribution of Grouped Data

Definition 1: A grouped frequency distribution is obtained by constructing classes (or intervals) for the data, and then listing the corresponding number of values (frequency count) in each interval.

Typically when the difference between the highest number and the lowest number is greater than 15, you might prefer to use a grouped frequency distribution.

The numbers are grouped into intervals, and the frequency of scores in each interval is listed in a separate column.

The intervals can be of any size, but, for ease of construction, a grouped frequency distribution should end up with no more than about 10 groups.

A grouped frequency distribution provides less precise information than does an ungrouped one, because **the individual performances are lost**. However, the benefit of a grouped frequency distribution is that one can understand any trends in the data at a quick glance

Basic Technical Terms
The following are some basic technical terms when a grouped frequency distribution is formed or data is classified according to class intervals.

a) Class limits:

The class limits are the least and the highest values that can be included in the class. For example, take the class 30-40.

The least value of the class is 30 and highest class is 40. The two boundaries of the class are known as the lower limit and the upper limit of the class.

The lower limit of a class is the value below which there can be no item in the class. The upper limit of a class is the value above which there can be no item to that class.

For the class 60-79, 60 is the lower limit and 79 is the upper limit, i.e. in this case there can be no value which is less than 60 or more than 79.

The way in which class limits are stated depends upon the nature of the data.

In statistical calculations, lower class limit is denoted by L and upper class limit by U.

b) Class Interval:

The class interval may be defined as the size of each group of data. For example, 50-75, 75-100, 100-125... are class intervals.

c) Width or size of the class interval:

The difference between the lower and upper class limits is called width or size of class interval and is denoted by 'C'.

d) Range:

The difference between largest and smallest value of the observation is called the range and is denoted by ' R' i.e. R = Largest value – Smallest value

R = L – S

e) Mid-value or mid-point:

The central point of a class interval is called the mid value or mid-point. It is found out by adding the upper and lower limits of a class and dividing the sum by 2.

Mid-value = $\dfrac{L+U}{2}$

For example, if the class interval is 20-30 then the mid-value is $\dfrac{20+30}{2} = 25$

f) Class Boundaries
Separate one class in a grouped frequency distribution from another. The boundaries have one more decimal place than the raw data and therefore do not appear in the data. There is no gap between the upper boundary of one class and the lower boundary of the next class. The lower class boundary is found by subtracting 0.5 units from the lower class limit and the upper class boundary is found by adding 0.5 units to the upper class limit.

Example 1: The number of unemployed people is given in the grouped frequency distribution table below :

Age Group	Number Unemployed
5-14	3,688
15-24	4,031
25-34	5,432
35-44	4,360
45-54	3,162
55-64	1,702

The class width = 14-5=9

The range= R = L – S=64-5=59

How to construct a grouped frequency distribution?

Step 1: Identify the highest value and the smallest value.

Step 2: Determine the class width. The class width is the difference between the two values in the given class.

All the time either the class width will be given or the number of classes

- If the number of classes is given:

$$\Rightarrow classwidth = \frac{highestvalue - leastvalue}{\#ofclasses}$$

- If the class width is given then we add it to the least value in the data set to obtain the upper limit of the class.

Step 3: When we start any class then the difference between the second number of the previous class and the first number in the coming class should be 1.

Remark: In the grouped frequency distribution, observe that the relative frequency column may not add up to exactly 1 and the percentage column may not add up to exactly 100 percent. This is due to the rounding of the relative frequency values.

Types of class intervals:

There are three methods of classifying the data according to class intervals:
a) Exclusive method
b) Inclusive method
c) Open-end classes

a) Exclusive method:

When the class intervals are so fixed that the upper limit of one class is the lower limit of the next class; it is known as the exclusive method of classification. The following data are classified on this basis.

Expenditure	No. of families
0 - 5000	60
5000-10000	95
10000-15000	122
15000-20000	83
20000-25000	40
Total	400

It is clear that the exclusive method ensures continuity of data as much as the upper limit of one class is the lower limit of the next class. In the above example, there are so families whose expenditure is between 0 and 4999.99. **A family whose expenditure is 5000 would be included in the class interval 5000-10000. This method is widely used in practice.**

b) Inclusive method:

In this method, the overlapping of the class intervals is avoided. Both the lower and upper limits are included in the class interval. This type of classification is used for a grouped frequency distribution for discrete variable like members in a family, number of workers in a factory etc., where the variable may take only integral values. It cannot be used with **fractional or negative** values like age, height, weight etc.

This method may be illustrated as follows:

Class interval	Frequency
5- 9	7
10-14	12
15-19	15
20-29	21
30-34	10
35-39	5
Total	70

Thus to decide whether to use the inclusive method or the exclusive method, it is important to determine whether the variable under observation is a **continuous or discrete one**. In case of continuous variables, the exclusive method must be used. The inclusive method should be used in case of discrete variable.

c) Open end classes:

A class limit is missing either at the lower end of the first class interval or at the upper end of the last class interval or both are not specified. The necessity of open end classes arises in a number of practical situations, particularly related to economic and medical data when there are few very high values or few very low values which are far apart from the majority of observations.

The example for the open-end classes is illustrated in the table:

Salary Range	No of workers
Below 2000	7
2000 – 4000	5
4000 – 6000	6
6000 – 8000	4
8000 and above	3

Example 2: A survey of 20 persons is taken .The following data is obtained. Each value represents the number of a certain product a person buys per month. Construct a frequency distribution using 5 classes.

```
14  12   26   10   15
28  25   20   17   18
35  24   19   34   35
22  30   33  27   39
```

Range = 39-10 = 29

Class width = 29/5 = 5.8 ≈ 6

Class limits	Class Boundaries	Frequency	Cumulative frequency
10 – 15	9.5 – 15.5	4	4
16 – 21	15.5 – 21.5	4	8
22 – 27	21.5 – 27.5	5	13
28 – 33	27.5 – 33.5	3	16
34 – 39	33.5 – 39.5	4	20