

**TABULATION AND GRAPHICAL REPRESENTATION OF DATA**

When a set of data contains only a few entries, a simple listing of the observations might be sufficient for interpreting the data. But when summarizing large masses of raw data, the simple listing of the observations may not be sufficient for the interpretation of the data. Here, it is often useful to distribute the data into groups or classes and presented in a table that gives the number of observations in each group.

**TABULAR REPRESENTATION OF DATA**

The data are usually organized into groups called classes and presented in a table which gives the frequency in each group. Such a frequency table gives a better overall view of the distribution of data and enables a person to rapidly comprehend important characteristics of the data.

**FREQUENCY TABLE (FREQUENCY DISTRIBUTION)**

A **frequency table** is a table that divides a data set into a suitable number of categories or classes. The term ‘frequency’ refers to the number of individuals or statistical unites in a category or class. We can also say that frequency is the number of times a particular value occurs in the set of data.

A **relative frequency distribution** is a table listing the possible values of the variable along with their relative frequencies (proportions).

**Frequency distribution for qualitative data**

A frequency distribution for qualitative data lists all categories and the number of elements that belong to each of the categories.

We are going to use the Tally Score Method to calculate frequency. This method entails making a stroke in the appropriate class for each observation and summing these for each class to obtain the frequency.

**Example**

The following list give the distribution of blood groups for 20 patients

**A AB O O B O B A O A O B A O B A B O AB O**

- Give a frequency distribution for these data.

Table 2.1: Frequency distribution of blood groups

Blood groups ( $x_i$ )	Tally	Frequency ( $n_i$ ) (Number of patients)
A		5
B		5
O		8
AB		2
$\Sigma$		20

**Relative frequency of a category**

The relative frequency of a category is obtained by dividing the frequency for a category by the sum of all the frequencies.

$$f_i = \frac{n_i}{\Sigma n_i}$$

**Percentage**

The percentage for a category is obtained by multiplying the relative frequency for that category by 100.

$$f_i\% = f_i * 100$$

Table 2.2: Relative frequency and peccentage distribution of blood groups

blood groups ( $x_i$ )	Relative frequency ( $f_i$ )	Percentage
A	$5/20 = 0.25$	$0.25 * 100 = 25\%$
B	$5/20 = 0.25$	$0.25 * 100 = 25\%$
O	$8/20 = 0.4$	$0.4 * 100 = 40\%$
AB	$2/20 = 0.1$	$0.1 * 100 = 10\%$

**Frequency distribution for quantitative data**

There are many similarities between frequency distributions for qualitative data and frequency distributions for quantitative data.

**Frequency distribution for ungrouped data (Single-valued classes):** An ungrouped frequency distribution is the one in which all the values are listed in an ascending or a descending order. If only a few unique values occur in a set of data, the classes are expressed as a single value rather than an interval of values. This typically occurs with discrete data but may also occur sometimes with continuous data.

**Example**

The table below gives the number of children per household in a sample of 30 households

5	4	3	4	2	2	5	4	4	2	5	4	2	3	2
3	5	4	3	5	4	3	5	4	5	4	3	4	5	3

- Make a tabular representation of the data

Table 2.3: Frequency distribution of children per household

Number of children ( $x_i$ )	Frequency ( $n_i$ )	Relative frequency ( $f_i$ )	Percentage
2	5	$5/30 = 0.1666$	$0.1666 * 100 = 16.66\%$
3	7	$7/30 = 0.2333$	$0.2333 * 100 = 23.33\%$
4	10	$10/30 = 0.3333$	$0.3333 * 100 = 33.33\%$
5	8	$8/30 = 0.2666$	$0.2666 * 100 = 26.66\%$
$\Sigma$	30		

**Frequency distribution for grouped data**

Here the classes are expressed as an interval of values. This typically occurs with continuous data. a number of steps are required, for the tabular presentation of data for grouped data. These steps are:

1. Determine the maximum and minimum values of the observations and **the range (R)**.

$$R = \text{maximum value} - \text{minimum value}$$

2. Determine the number of classes: we can determinate the number of classes (K) by the following rules - **Sturges rule:**

$$K = 1 + 3.322 \log (n)$$

Where: k: the number of classes, n : the number of observations, log ( ) : the decimal logarithm

- **Yule’s rule:**

$$K = 2.5 \sqrt[4]{n}$$

The number of classes should be between 5 and 15

3. Determine the **class width (class size, or class length)** for the distribution as follow:

$$\text{class width (L)} = \frac{\text{The Range}}{\text{number of classes}} = \frac{R}{K}$$

4. Determine the class intervals: Each class has a lower class limit and an upper class limit.

A class interval that, has either no upper class limit or no lower class limit indicated is called **an open class interval**.

The difference between the boundaries for any class gives the class width (L).

5. Find **the class mark -  $C_i$  - (the class midpoint)** : The class mark for a class is the midpoint of the class, it can be obtained by adding the lower class limit ( $L_{iL}$ ) to the upper class limit ( $L_{iU}$ ) for the class and the sum divided by 2.

$$C_i = \frac{L_{iL} + L_{iU}}{2}$$

6. State the number of observations falling into each class interval

**Cumulative frequency distributions**

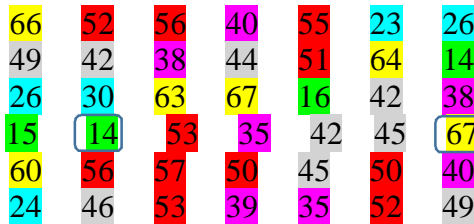
A cumulative frequency distribution gives the total number of values that fall below various class boundaries of a frequency distribution. Sometimes it is useful to consider a cumulative-frequency distribution of all values greater than or equal to the lower class boundary of each class interval (or all values less than the upper class boundary of each class interval). This is sometimes called an **“or more” cumulative distribution**, while the one considered above is a **“less than” cumulative distribution**.

**Cumulative relative frequency distributions**

A cumulative relative frequency is obtained by dividing a cumulative frequency by the total number of observations in the data set.

**Example**

Monthly wages sample (in thousands of dinars) for a company’s workers are listed.



- Make a frequency distribution of the data set. Include class limits, midpoints, frequencies, relative frequencies, and cumulative frequencies.

**Solution :**

(a) Frequency distribution table:

- Minimum values of the observations = 14
- Maximum values of the observations = 67
- The range:  $R = 67 - 14 = 53$
- The number of classes: we can determinate the number of classes by following Sturges rule  $K = 1 + 3.322 \log (42) = 6.39 \approx 6$
- Class width:  $L = \frac{53}{6} = 8.83 \approx 9$

Table 2.4: Frequency distribution of monthly wages

Class Range	Frequency	midpoints	Relative frequency	Cumulative frequency less than	Cumulative frequency or more	Cumulative relative frequency
[14 23[	4	18.5	0.095	4	42	0.095
[23 32[	5	27.5	0.119	9	38	0.214
[32 41[	7	36.5	0.167	16	33	0.381
[41 50[	9	45.5	0.214	25	26	0.595
[50 59[	11	54.5	0.262	36	17	0.857
[59 68[	6	63.5	0.143	42	6	1
$\Sigma$	42	/	1	/	/	/