Graphic Presentation of Data

5

5.1 Construction of Graph

1. In the construction of graph, two simple lines are first drawn, which cut each other at right angles. These are called axis. The horizontal line is called "abscissa" or x-axis, and the vertical line is called "ordinate" or y-axis. The point of intersection is called the point of origin. The basic construction of graph has been depicted in Fig. 5.1.
2. Draw a graph for practice keeping in mind the values of x and y as illustrated above.

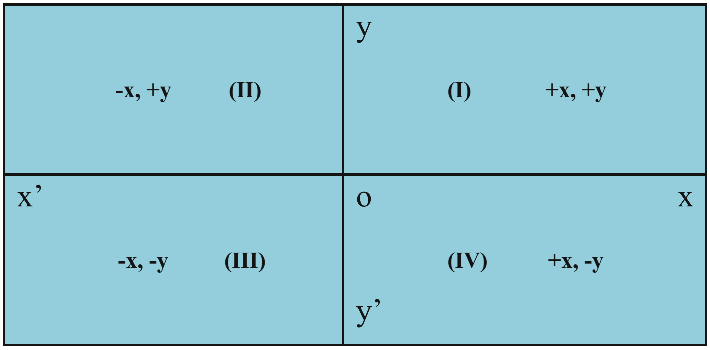

Fig. 5.1 In the above figure, x'ox is the abscissa and yoy' is the ordinate. "o" is the point of origin. In the quadrant "I," values of x and y are positive. In the quadrant "II," values of x are negative, and values of y are positive. In the quadrant "III," values of x and y are negative. In the quadrant "IV," values of x are positive, and values of y are negative

C. S. Rayat, *Statistical Methods in Medical Research*,
https://doi.org/10.1007/978-981-13-0827-7_5

5.2 Choice of Scale

1. The graph should be condensed in the space provided, choosing the appropriate scale for x-axis and y-axis. Examining the magnitude and sign of values, one can determine the positions of x-axis and y-axis.
2. Generally an independent variable is shown on the x-axis and a dependent variable on the y-axis. The horizontal scale need not begin with zero, but the vertical scale must begin with zero. If the fluctuations in the values of a variable presented on the y-axis are very small as compared to the size of the item, a false baseline is used.
3. In the natural scale, equal space represents the equal amount of magnitude. There is no definite relationship between the lengths of "abscissa" and the length of the "ordinate." It is a convention that x-axis is taken 1.5 times of the length of y-axis, but there is no strict rule.

5.3 Plotting of Data

1. After the scales have been decided and marked on the graph paper, the data can be plotted. If the variable is a continuous one, these points may be joined to give a smooth curve. If the data relates to the discrete variable, the points may be joined by straight lines.
2. Sometimes it is very difficult to smooth curves even in case of continuous variable, so the data may be shown by joining the points with straight lines. However, curves obtained by mathematical relationships must be smoothed, and these should not be shown by joining with straight lines.

5.4 Graphs of Time Series or Historigrams

1. Line charts are used to present the historical data due to their superiority over "bar charts." Line charts show the continuity of the variables, whereas "bar charts" indicate the discontinuity of the variables.
2. More time is consumed in depicting the data in the form of bars than in the form of "line chart." Line charts give quick understanding about the movement and the absolute magnitude of variable. Secondly, it is possible to interpolate a value from the graph, for the year for which the data was not available.
3. The graph shows the changes in the values of a variable. If the absolute values are taken into consideration, the graph drawn is known as "absolute historigram." If the graph is plotted from the index numbers, it is then called "index historigram."

5.5 Comparison of Time Series

1. If two or more variables, expressed in the same unit of measurement, are to be compared, these may be presented on the same graph choosing the scales suitable for all the variables. The same variable at two or more periods from the same place may also be represented on the same graph for comparison.
2. In order to compare the changes between different series, it is necessary to have a common base year. Equating the frequency of base year to 100 in each series, the percentage of other frequencies can be worked out. These percentages of each series may be plotted and the lines drawn to show the changes taking place. Thus, the relative changes of all the series can be compared.
3. Sometimes “mixed graphs” are prepared to study interrelated variables. In such graphs, one variable is usually shown in a “bar diagram” and the other in the shape of a “line diagram” or curve.

5.6 Semilogarithmic Scale

1. In case of natural scale, the absolute changes in a variable are studied. But to study relative changes of two or more variables or to study rate of change in a variable, a semilogarithmic scale is appropriate. In this case, y-axis is scaled logarithmically, whereas x-axis is in the linear scale. In the semilogarithmic scale (ratio scale), the length of the interval between two values is proportional to the ratio between these two values. Hence, equal spaces in the ratio scale represent equal logarithmic differences or equal ratios.
2. When to use semilogarithmic scale: (a) when comparison between series of widely different magnitudes is desired, (b) when comparison between series of different units is required, and (c) when relative change is to be studied or compared with various series.

5.7 Interpretations of Semilog Curves

1. If a semilog curve rises upward, it indicates that the growth is positive and the values are increasing.
2. If the curve is a straight line and is ascending, it indicates that the curve is increasing with a constant rate of growth.
3. If the curve is a straight line and is descending, it indicates that the curve is decreasing with a constant rate.
4. If the curve rises more strongly at one point and then at another, it indicates that the rate of growth is more in the former case than in the latter case.
5. If two curves are parallel to each other, their rate of increase or decrease remains the same.
6. If one curve is steeper than the other, its rate of change is higher than the other.

5.8 Properties of a Logarithmic Scale

1. Equal distances on the vertical scale represent equal proportionate changes.
2. Logarithmic scale does not begin with zero.
3. It is not possible to study an aggregate in its component parts.
4. Logarithmic scale cannot be used for studying absolute changes.

5.9 Normal Frequency Curve: Properties of Normal Frequency Curve Are Given Below

1. It is a unimodal, perfectly symmetrical bell-shaped curve.
2. Frequency decreases on either sides of the central value.
3. Frequency at the central value is the highest.
4. Mean, median, and mode coincide with the central value.
5. Frequencies of this distribution are in a definite mathematical relationship.
6. Total area under the normal frequency curve is equal to the total number of observations.
7. Number of observations falling between any two ordinates could be estimated.
8. Areas between "population mean" (μ) and "standard deviation" (σ) have been depicted in "tabulated" and "graphic" form in Table 5.1 and Fig. 5.2, respectively.
9. Near about the mean value, the curve is convex toward x-axis, whereas near the two tails, it is concave.
10. The points of inflection (the points where the change in curvature occurs) are at a distance of $\pm 1\sigma$, $\pm 2\sigma$, and $\pm\ 3\sigma$ on either side of the mean value (μ).
11. The values of all odd moments about the mean are zero, so the "skewness" is zero.

Table 5.1 Area between "population mean" (μ) and "standard deviation" (σ)

Area of "population mean ± SD"	Percentage (%) of observations
$\mu \pm \sigma$	68.27
$\mu \pm 1.96\ \sigma$	95.00
$\mu \pm 2\ \sigma$	95.45
$\mu \pm 2.576\ \sigma$	99.00
$\mu \pm 3\ \sigma$	99.73

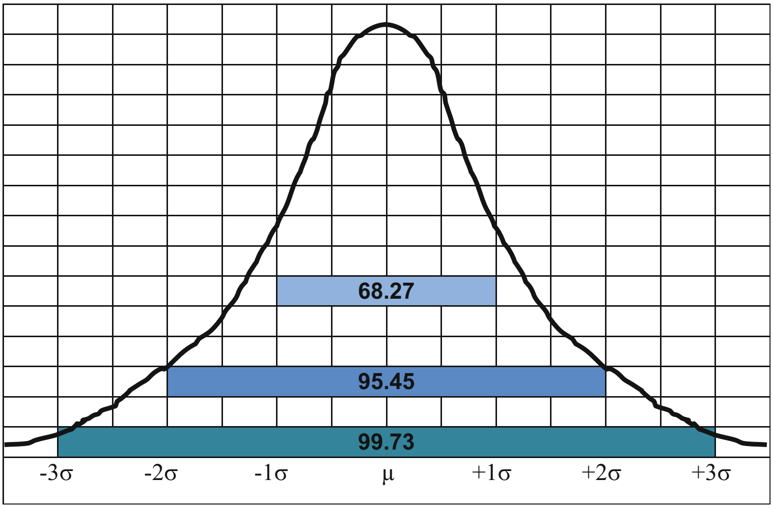

Fig. 5.2 Normal distribution graph showing percentage of observations on either side of "population mean" (μ) in relation to "SD" (σ)

5.10 Moderately Asymmetrical Curve

1. Curve is not symmetrical.
2. The distribution of frequencies in such a curve is not in a mathematical relationship.
3. Curve may either be "positively skewed" or "negatively skewed."

5.11 Extremely Asymmetrical Curve

1. The "skewness" is very high in this type curve.
2. The cluster of observations having maximum frequency is generally on one corner and not in the middle as in the case of symmetrical curve.