1. Introduction

Welcome to the series of e-Learning modules on histogram. In this module we are going to cover the meaning, utility, construction, analysis and interpretation of a histogram.

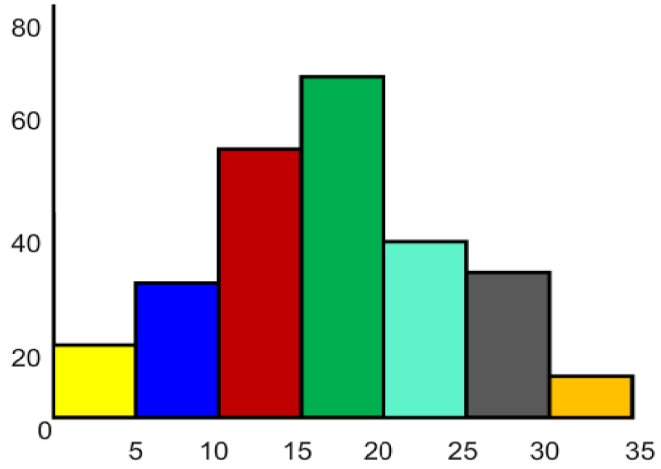
At the end of this session, you will be able to:

- Explain the meaning and use of histogram
- Explain how to do the construction of a histogram
- Explain how to do the analysis and interpretation of a histogram

What is a Histogram?
A histogram is a two dimensional area diagram as it represents both the class interval and the frequency of the data. It is the most popular and widely used graphical representation of the data as a set of vertical adjacent rectangle bars.

A histogram is a graphical display of data using bars of different heights. Total area of the rectangles in a histogram represents the total frequency. A histogram with more number of class intervals is more effective in depicting the structure of the frequency distribution.

Figure 1



Definition:
According to Opermann -“A histogram is a bar chart or graph showing the frequency of occurrence of each value of the variable being analyzed.”

2. Uses and Parts of a Histogram

Let us discuss some of the uses of histogram.

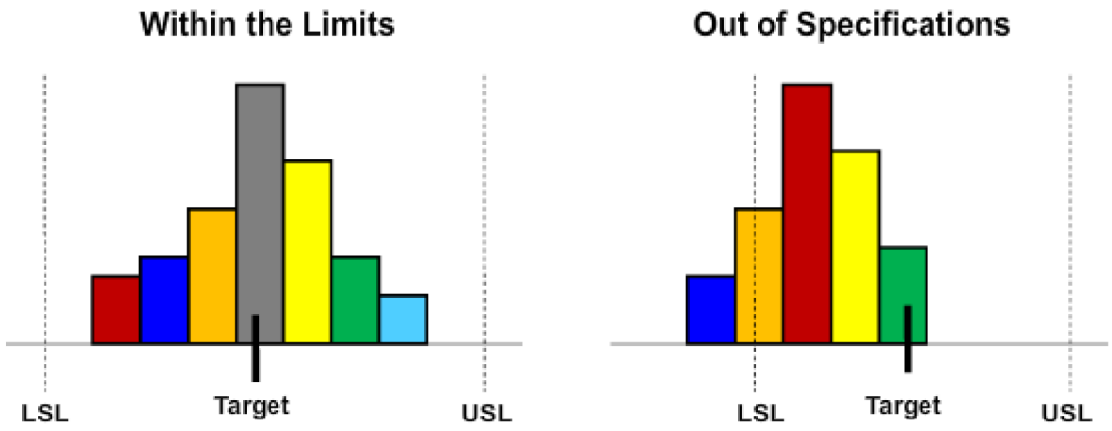
- Histogram is a popular tool: A histogram is a popular tool to organize and display a large set of measurements presented in a table, into a more user friendly format
- Easy to understand: A Histogram helps the viewer to understand where the majority of values falls in a measurement scale, and how much variation is there in the data

Summarizing large data sets graphically: You can make it much easier to understand by summarizing the set of data presented in a table on a tally sheet and organizing it into a histogram

Compare process results with specification limits: Specification limits may take the form of length, weight, density, quantity of materials to be delivered, or whatever is important for the variable. Histogram is a graph on which the specific limits are mentioned and helps to determine quickly whether the variable is able to produce "good" results or not

Here, this histogram graph shows the Specification within the limit and out of specification limits.

Figure 2

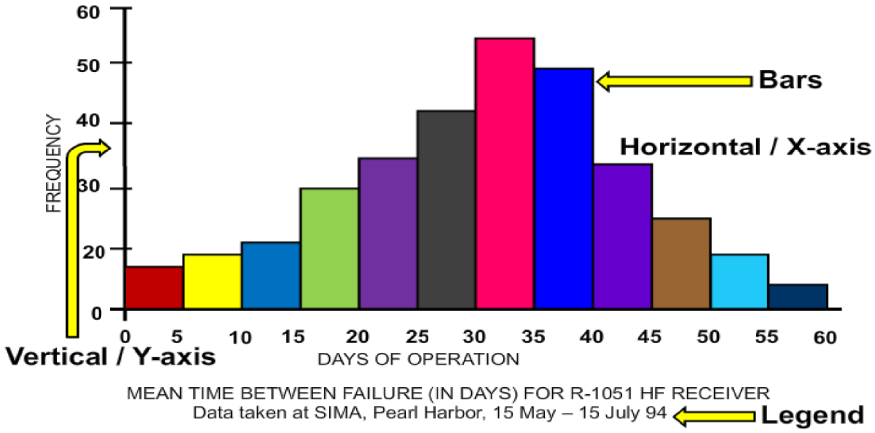


Communicate information graphically: Histogram helps in representing large sets of data with specific limits which helps the viewer to easily compare the values which occur most frequently and communicate the information in a simple manner

Helps in decision making: Histogram is a graph with specific shape, size and the spreading of the data. If these data are authentic and has been collected from the right source then it will help in making decisions or predictions easily. If the data you have is not of recent time period and its collection if not authentic then the usage of a histogram chart is a waste. Measurements cannot be used for making decisions or predictions when they were produced by a process that is different from the current one, or were collected under unknown conditions

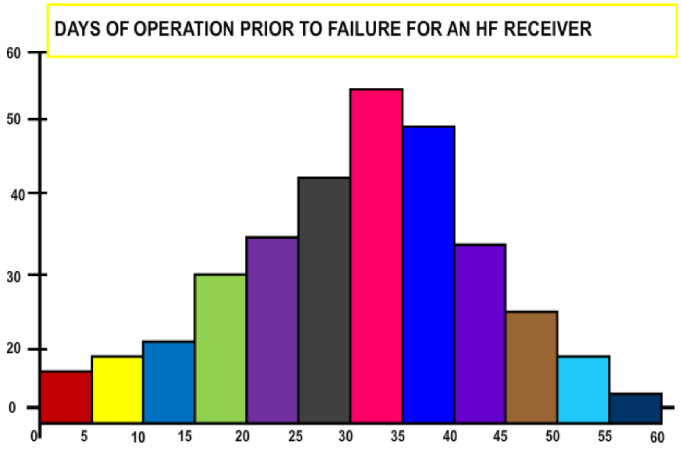
Now we will discuss what are the parts of histogram.

Figure 3



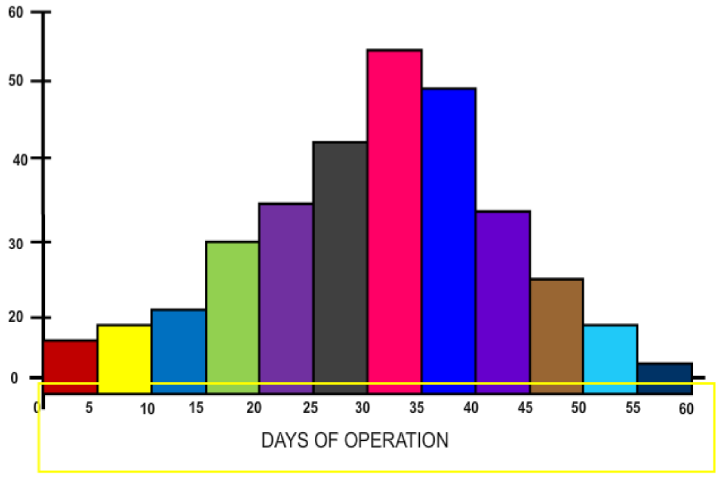
1. Title:
The title briefly describes the information that is contained in the Histogram.

Figure 4



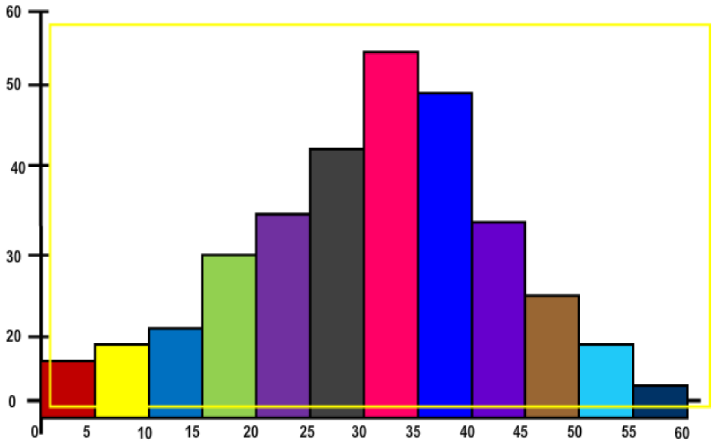
2. Horizontal or X-Axis:
The horizontal or X-axis shows you the scale of values into which the measurements fit. These measurements are generally grouped into intervals to help you summarize large data sets. Individual data points are not displayed.

Figure 5



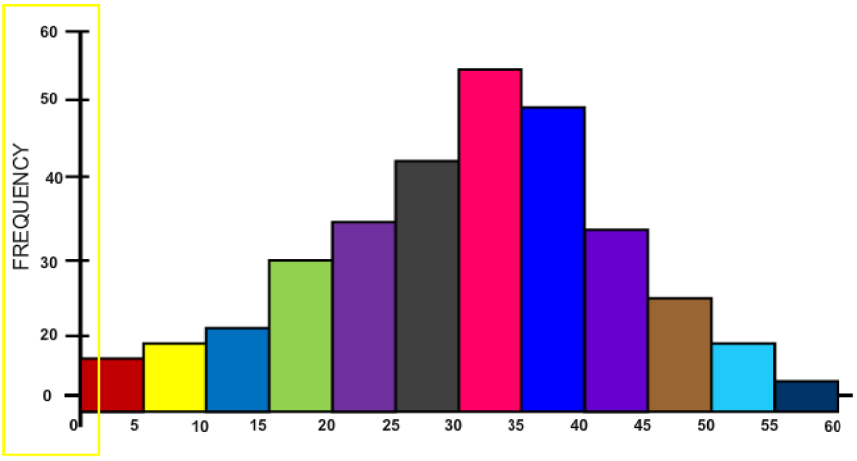
3. Bars:
The bars have two important characteristics—height and width. The height represents the number of times the values within an interval occurred. The width represents the length of the interval covered by the bar. It is the same for all bars.

Figure 6



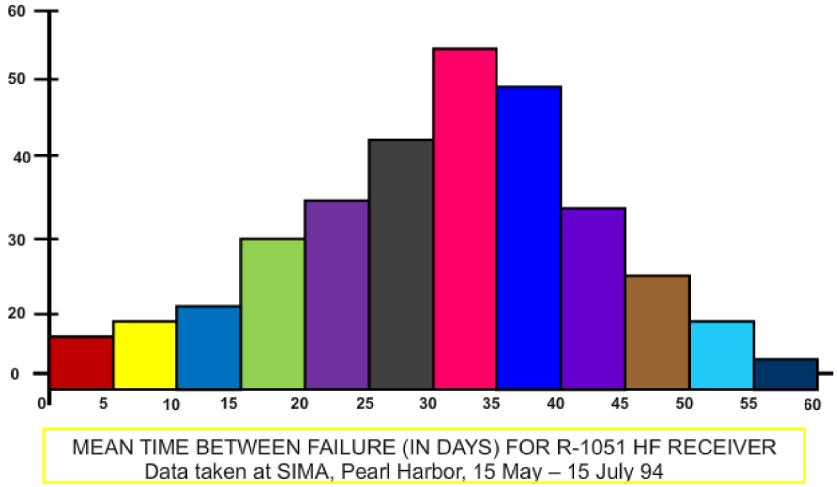
4. Vertical or Y-Axis:
The vertical or Y-axis is the scale that shows you the number of times the values within an interval occurred. The number of times is also referred to as "frequency."

Figure 7



5. Legend:
The legend provides additional information like, where the data came from and how the measurements were gathered.

Figure 8



3. Construction of a Histogram

The construction of a histogram from a given data is constructed using the following steps:
Step 1:
Calculating the data points: In this step we try to count the data which is in the raw form. It is a form of data where all the values are shown separately and is called an Individual series. In this form the data are disorganized and give us only the count of the data points collected for the sake of analysis. Thus, in this step we simply count the total number of entries we have made to complete the data set.

Step 2:
Organizing the data points: In this step we summarize the data using a simple mathematical sorting technique to arrange the data in Ascending Order- where the data is arranged serially starting from a small value to a big value. It is easy for interpretation. For this purpose the collected data is rearrange in a table with two columns one representing all possible values of the variable and the other representing Tally bar for every occurrence of the number in the series. This step helps in condensing the data points.

Step 3:
Calculating the class Intervals: In this step we have used a simple technique to group the data points in a range - starting from the lowest value, called “Class Intervals.” Class Intervals can also be obtained by subtracting the largest and the smallest value of the data.

Step 4:
Determine the number of intervals: In this step we try to identify the number of intervals needed as this helps in designing the pattern, shape or spread of the histogram. In simple it helps in deciding the number of rectangle bars needed in a graph.
Generally if there are less than 50 data points we will have 5 to 7 intervals, 50 – 100 data points we will have 10 intervals, 100 – 250 we will have around 15 intervals and more than 250 we will have around 20 intervals.

Step 5:
Computing the width of the class intervals: To compute the interval widths divide the range by the number of intervals. When computing the interval width, you should round the data up to the next higher whole number to come up with values that are convenient to use.

For example, if the range of data is 17, and you have decided to use 9 intervals, then your interval width is 1.88. You can round this up to 2.

Step 6:
Determine the starting point of the class interval: Use the smallest data point in your measurements as the starting point of the first interval. The starting point for the second interval is the sum of the smallest data point and the interval width

Step 7:
Calculating the number of data in each class interval: Count the number of points that fall within each interval. These are the data points that are equal to or greater than the starting

value and less than the ending value. It is the data points greater that the beginning of the interval and lesser than the end of the interval. It has to be noted that each data point will appear only in one class interval.

Step 8:
Plotting the data in the graph: In this step we plot the data in a graph keeping in mind that the class intervals will appear in the horizontal axis, the frequencies in the vertical axis, and the height of the rectangle bars representing the frequency of each class interval. Check whether the graph so performed is logical and reasonable.

Step 9:
Add title and legend: In this step we try to give a title and legend to the graph so that it helps in identifying the five W of the data (Who, Where, What, When and Why) that are important for understanding and interpreting the data. The legend can have information like the sample size or the time period or the source of the data collected for easy clarification or description of the information of the data.

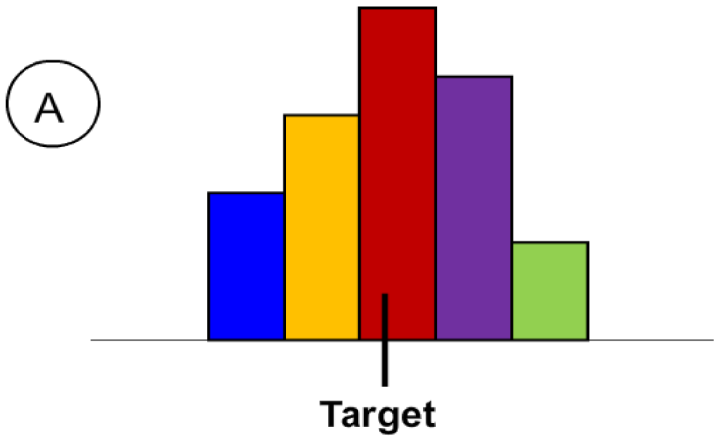
4. Interpretation of a Histogram

After knowing how to construct a histogram we will now see how to interpret a histogram.

A Histogram is a visual representation which gives an idea of the location of the various measurements and the kind of variation the data has.

A Histogram can show any of the following conditions.
• Maximum data can be closer to the target with a little variation in it. (Diagram A).
This plot represents data with a well-defined peak that is close in value to the median and the mean (TARGET). While there are "outer layers," they are of relatively low frequency. Thus, it can be said that deviations in this data group from the mean are of low frequency.

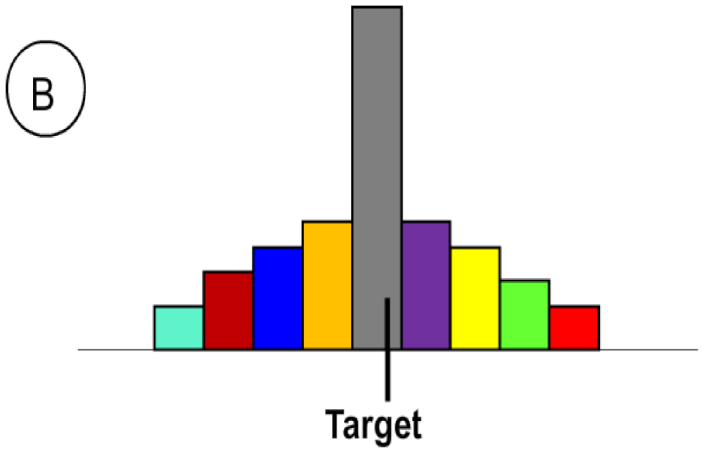
Figure 9



- Sometime a reasonable part of the data points can be on the target and many others dispersed away from it. (Diagram B)

In this plot the peak is still fairly close to the median and the mean (TARGET) but it is much less defined. There are almost as many values close to the peak as at the peak itself and outer layers are frequent.

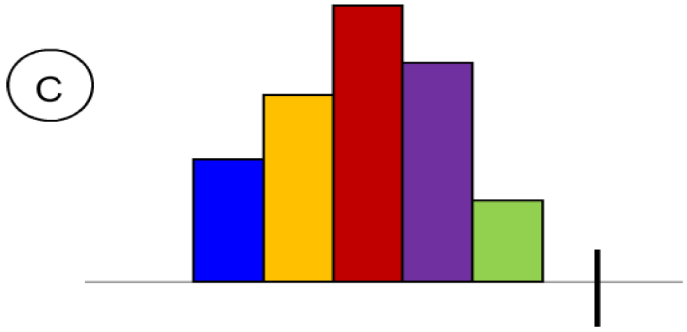
Figure 10



- There can be a situation where most of the data could be closely related but located away from the target. (Diagram C)

In this plot the peak is still fairly away from the median and the mean (target) but it is much less defined. We can tell from the plot that the location of the peak is closer to the values but still the outer layers are away from the peak.

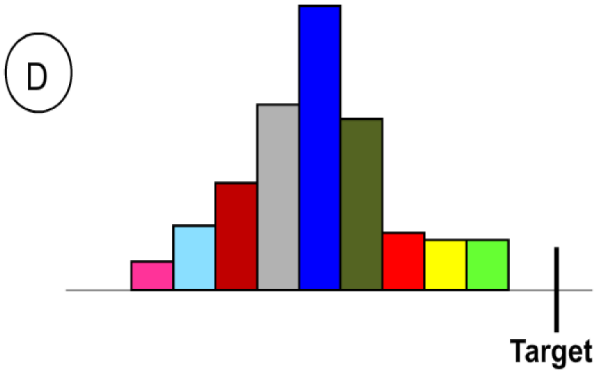
Figure 11



- The data could be away from the target and widely dispersed. (Diagram D)

In this plot the peak is still fairly away to the median and the mean (TARGET) but it is much less defined. There are almost as many values close to the peak as at the peak itself and outer layers are widely dispersed.

Figure 12



The benefits of interpretation of the data by histogram helps in:

- Understanding the consistency of the variable
- Significant findings that can contribute for decision making
- Identifying the performance of the variable within the specific limits
- Identifying the relationship of the variable with the achievement target
- Portraying the defects of the variable
- Predict current and future course of action
- Defining the stability of the variable and its process
- Identifying fundamental changes to eliminate causes in variation

Conclusion:
The Histogram may not be symmetrical, the shape may show that something is wrong, the data from several sources are mixed, different measurement devices were used, or operational definitions weren't applied. What is really important here is to avoid jumping to conclusions without properly examining the alternatives.

5. Illustration for the Construction of a Histogram

Illustration:
Let us take an example to understand in detail the construction of a histogram:

Following is the market survey response of people with regard to the acceptability of a new product in the market.

Table 1

40	45	41	45	30	39	8	48	25	45
26	11	23	24	13	29	8	40	41	42
39	35	18	25	35	40	42	43	44	36
37	32	28	27	25	26	38	37	40	35
32	28	40	41	43	44	45	40	39	41

Let us construct a histogram for the above data by applying the nine steps we have discussed in this module. Let us also interpret the graph to understand the behavior of the respondent in acceptability of the new product.

Step 1 – Calculating the data point:
Here, the total number of data point is 50.

Table 2

40	45	41	45	30	39	8	48	25	45
26	11	23	24	13	29	8	40	41	42
39	35	18	25	35	40	42	43	44	36
37	32	28	27	25	26	38	37	40	35
32	28	40	41	43	44	45	40	39	41

Total = 50 (41)

Step 2 - Organizing the data point:
Here, the collected data is rearranged in a table of two columns- one representing all possible values of the variable in Ascending Order and the other representing tally bar for every occurrence of the number of series.

Step 3 - Calculating the number of intervals:

Largest value in this data is 48 and the smallest value in the data is 8 therefore the range of the data is 48 minus 8 is equal to 40.

Step 4 – Determine the number of intervals:
As the number of data points in this data set is 50 let us take 5 class intervals.

Table 3

If there are this many data points	This intervals are used
Less than 50	5 - 7
50 – 100	10
100 – 250	15
More than 250	20

Step 5 – Computing the width of the class interval:
Interval width is equal to range divided by the number of intervals which in this case is 40 divided by 5 is equal to eight. Therefore, the width of the class interval in this example is equal to eight.

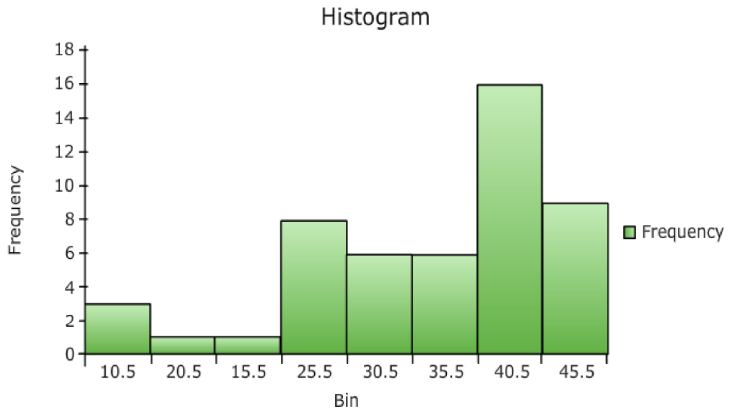
Step 6- Determining the starting point of the class interval:
Once we have calculated the range and the width of the class interval. We will determine the starting point of the first interval as the smallest value of the data set add the value of the interval width to arrive at the ending point.

Step 7 - Calculating the number of data points in each class interval:
In this step we will add up the data points that will appear between the starting point and the ending point. This will help in determining the frequency of each class.

The following table represents the data according to step 6 and 7 denoting the starting point class interval width and the number of data points in each class intervals of the data set.

Step 8 - Plotting the data in the graph:
In this step we will represent the class intervals the variables of the data set that is the acceptability level of the respondents in the X-axis and the frequency that is the number of respondents in each class interval in the Y- axis.

Figure 13



Step 9 - Add title and legend:
In this step we will add the title and legend to the plotted graph. The title for this graph will be acceptability of the new product and the legend could be time and period of data collected.

Now we will discuss about the advantages and disadvantages of Histogram.

Advantages of Histogram are:

- Each rectangle shows distinctly separate class in the distribution
- The area of the rectangle in relation to all other rectangles shows the proportion of the total number of observations pertaining to that class

Disadvantages of Histogram are:

- It cannot be constructed for open ended classes
- It can be misleading if the distribution has unequal class intervals and suitable adjustments in frequencies are not made

Difference between Histogram and Bar Graph:

Table 4

Histogram	Bar Graph
It consists of rectangles touching each other.	It consists of rectangles normally separated from each other with equal space.
The frequency is represented by the area of each rectangle.	The frequency is represented by height. The width has no significance.
It is two dimensional.	It is one dimensional. It is used as a virtual aid to represent data.

Here's a summary of our learning in this session:

- The meaning and use of histogram
- Construction of a histogram
- Analysis and interpretation of a histogram