

Histogram

What is it?

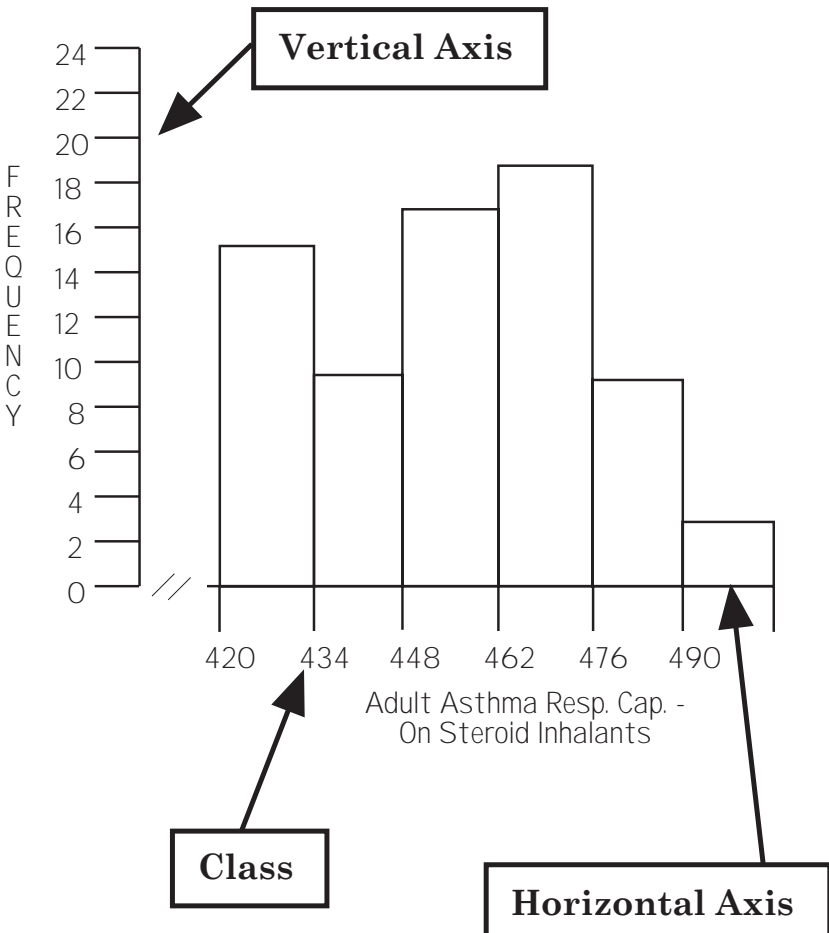
A histogram is a bar graph representing the frequency of individual occurrences or classes of data. A histogram shows basic information about the data set, such as central location (mean, median, and mode), width of spread (range or standard deviation), and the shape.

The purpose of making a histogram is to gain knowledge about the system. This knowledge, gained from the basic information given by the histogram (central location, spread, and shape), will act as a guide to improve the system. From a stable system, predictions can be made about the future performance of the system. If the system were unstable, it would change from time to time and the histogram would have little predictive value.

The group uses a histogram to assess the system's current situation and to study results. The histogram's shape and statistical information help us know how to improve the system. After an improvement action is carried out, the group continues to collect data and make histograms to see if the theory has worked.

What does it look like?

A completed histogram is shown below. An outpatient clinic patient health educator constructed this histogram using data from the \bar{X} -R chart for the Adult Asthmatic Patient Respiratory Capability. The \bar{X} -R chart showed the system to be unstable. The patient and care provider successfully identified special cause in the last four subgroups (the patient was out of town and forgot to take medications). Deleting the four subgroups occurring due to special cause, the educator used the remaining 23 subgroups to make this histogram. (See Step 9 in \bar{X} -R for the stable control chart using the 23 subgroups.)



When is it used?

Use a histogram when you can answer “yes” to both these questions:

- 1. Do you have a data set of related values, either attributes (counts) or variables data (measurement)?** For analyzing system performance, single readings or individual data points are of limited value. Much more can be learned from a group of data points because they reflect the system’s variation. Using a histogram is one way to start learning from a group of data points.
- 2. Is it important to visualize central location, shape, and spread of the data?** When it comes to data analysis, a picture is worth a thousand words. Seeing the form of the data makes it easier to understand the kind or pattern of variation the system is producing.

How is it made?

These steps assume that the data for the construction of the histogram has already been collected. The data can be collected especially to make a histogram or can come from the data entry section of a control chart. Once you have collected data for a control chart, that same data could be used to make a histogram. The data entry section of the control chart used for the example histogram is shown below.

VARIABLES CONTROL CHART

VARIABLES CONTROL CHART													Quality Measure		Chart No.									
Product / Service					Asthma Care					Process			Respiratory Process			Specification Limits		N/A						
User Name		SKM			Location			Home			Measurement Device			Home Spirometer			Unit of Measure		MOSM					
DATE	5-17	5-18	5-19	5-20	5-21	5-22	5-23	5-24	5-25	5-26	5-27	5-28	5-29	5-30	5-31	6-1	6-2	6-3	6-4	6-5	6-6	6-7	6-8	
SAMPLE MEASUREMENT	1	430	460	450	475	440	480	420	480	450	430	470	475	480	500	450	465	460	445	430	450	500	420	420
	2	420	480	470	470	450	450	460	480	470	450	445	480	450	450	430	470	470	450	450	470	440	440	430
	3	440	470	470	485	460	465	430	470	465	440	440	470	470	470	430	480	485	430	430	470	430	450	455
	4																							
	5																							
SUM		1790	1410	1390	1430	1350	1395	1310	1420	1385	1320	1355	1425	1400	1540	1310	1415	1415	1325	1310	1390	1370	1310	1305
AVERAGE \bar{x}		423	470	463.3	476.7	450.0	465.5	436.4	476.7	461.7	440.0	451.7	475.0	466.7	473.3	436.7	471.7	471.7	441.7	436.7	463.3	456.7	436.7	435.8
RANGE R		20	20	20	15	20	30	40	10	20	20	30	10	30	50	20	15	25	20	20	20	70	30	35
NOTES																								

- 1. Select the classes.**
 - a. Determine the number of classes.**

To find the number of classes (or subdivisions) needed for the histogram, first count the number of data points in the data set. Then use the following table to choose the number of classes. As the table indicates, it is best to use no fewer than 5 classes (or subdivisions) or more than 20.

<u>No. of Data Points</u>	<u>No. of Classes</u>
Under 50	5 - 7
50 - 100	6 - 10
100 - 250	7 - 12
Over 250	10 - 20

There are 69 data points in the example, 23 subgroups of 3 observations each. This table indicates between six and ten classes should be used for this many data points. Choose 6 for the example. The choice of the number of classes you want to use is only a rough estimate at this point. You can decide later to use more or fewer classes.

b. Determine the class width and boundaries.

The width of the class determines the range of data points in each class. Find the class width by dividing the range of the data set by the number of classes (found in Step a). The range is found by subtracting the smallest value in the data set from the largest.

$$Range = X_{highest} - X_{lowest}$$

In this example, the highest value in the data set is 500 and the lowest is 420. So the range is:

$$\begin{aligned} Range &= 500 - 420 \\ &= 80 \end{aligned}$$

The class width for the example is:

$$\begin{aligned} Class\ Width &= \frac{range\ of\ data\ set}{No.\ of\ classes} \\ &= \frac{80}{6} \\ &= 13.3 \\ &= 14 \end{aligned}$$

Round the class width to an easy number to work with. In the example, we rounded 13.33 to 14.

Next, select a starting number for the lower boundary of the first class. The lower boundary should be chosen so the lowest value in the data set is included in the first class. A convenient lower boundary for the example is 420, since the lowest value in the data set is 420.

To determine the lower boundaries for the remaining classes, begin with the lower boundary of the first class and add the class width. Continue adding class width until the number of classes is complete and all the data has been included.

The lower class boundaries for this example are:

$$420 + 14 = 434$$

$$434 + 14 = 448$$

$$448 + 14 = 462$$

$$462 + 14 = 476$$

$$476 + 14 = 490$$

$$490 + 14 = 504$$

In some cases, an extra class may need to be added so the highest data point will be included.

The upper boundary for each class is any number under or below the lower class boundary of the next class. For example, the upper class boundary for the first class is “under 434.” This means that any number greater than or equal to 420 but less than 434 falls into the first class. This is done so that no point will fall on the boundary between two classes.

The classes for the example are:

420 to under 434
434 to under 448
448 to under 462
462 to under 476
476 to under 490
490 to under 504

2. Record the data.

The easiest way to record the data is to create a check sheet listing the classes along the left side with space to the right to make tally marks. To record the data, make a tally mark beside the class in which each data point falls. Total the number of marks in each class. Shown below is the completed check sheet for the example.

CLASSES	TALLY	TOTAL
420 UNDER 434		13
434 UNDER 448		9
448 UNDER 462		17
462 UNDER 476		19
476 UNDER 490		9
490 UNDER 504		2