Online library of Quality, Service Improvement and Redesign tools

# Histogram

NHS England and NHS Improvement

# Histogram

## What is it?

A histogram is a type of bar chart that displays a set of continuous data that you can use to evaluate the distribution or variation of data over a range, for example, weight, length of time, size and age.

The shape of the bar or curve shows the distribution of data; it can help you determine what statistical tests to apply and understand where problems are. With this information, you can decide where to focus your improvement efforts.

## When to use it

A histogram is a good tool for early analysis in a project. It is useful when you have collected some baseline data and want to understand where to focus your improvement efforts in a patient pathway. For example, you may wish to improve flow of high volume elective procedures.

If you have created **run charts**, you can use the data, as both can be used to show a before and after picture of the process under review. A histogram does not show you how things are changing over time – a run chart will give you this information.

## How to use it

1. Gather data on a process you are interested in – it could be time, weight, size or frequency of occurrences. To effectively see the patterns in performance, try to collect at least 50 data points. A quick way to generate lots of ideas is to ask 20 people to suggest five each.
2. To 'draw a picture' of the set of data, you need to split the data points into several non-overlapping classes of equal width. Decide how many bars (or classes) you want in your histogram. Table one suggests how many classes you could use based on the number of data points you have collected. Generally speaking, if a histogram has too many classes because the class width is too small, we can lose sight of the overall shape of the distribution. On the other hand, if there are very few classes because the class width is too large, the data are lumped together and again, the shape of the distribution is missed. A good rule of thumb when constructing a histogram is to use between 5 and 20 classes.

**Figure 1: Table of data**

| Number of data points | Number of classes |
|---|---|
| Under 50 | 5–7 |
| 50–100 | 6–10 |
| 100–250 | 7–12 |
| Over 250 | 10–20 |

3. Identify how many of your data points fall into each class and record this information in a table. This will be your frequency table.
4. To construct the histogram, draw a two axis graph. Classes are plotted on the x-axis (horizontal) and their frequency is plotted on the y-axis (vertical). Transfer the data from your frequency table onto the graph, so that you have a vertical bar for each class.
5. Interpret your results.

## Example

A team was thinking about how its directorate could improve the quality and value of the services provided. They had a hunch that the length of stay varied greatly, so decided to collect some baseline data.

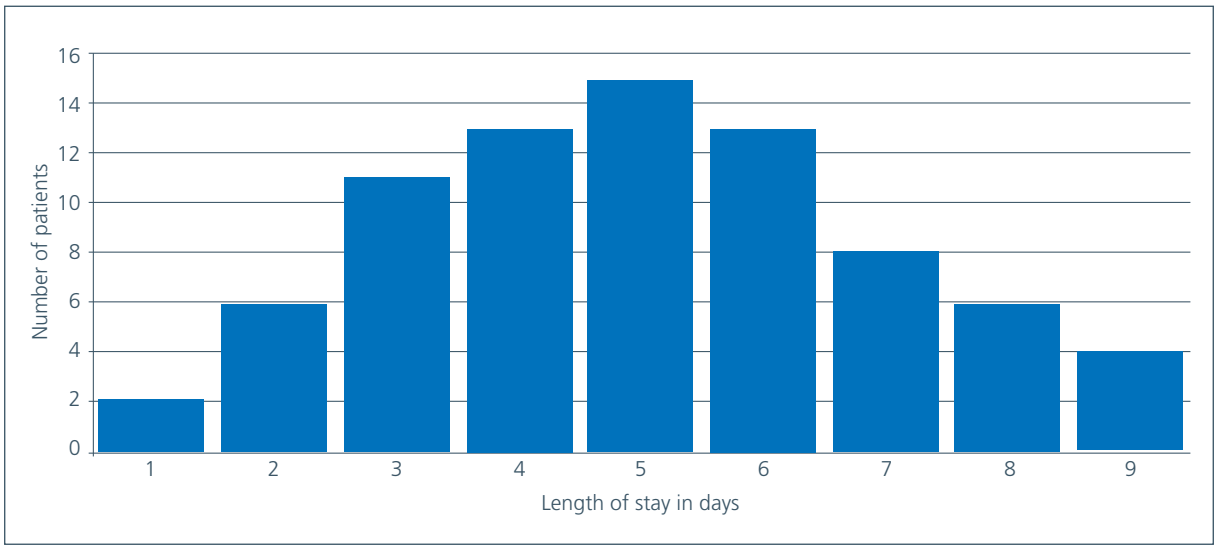
They collected length of stay data from one ward for a month, which they turned into a frequency table:

**Figure 2: Frequency table**

| Length of stay (days) | Number of patients |
|---|---|
| 1 | 2 |
| 2 | 6 |
| 3 | 11 |
| 4 | 13 |
| 5 | 15 |
| 6 | 13 |
| 7 | 8 |
| 8 | 6 |
| 9 | 4 |

The team then converted this into a histogram:

**Figure 3: Histogram 1**




## Interpreting the results

The data in this example shows a normal distribution. This is characterised by a bell-shaped curve with data more or less symmetrical about the central mean.

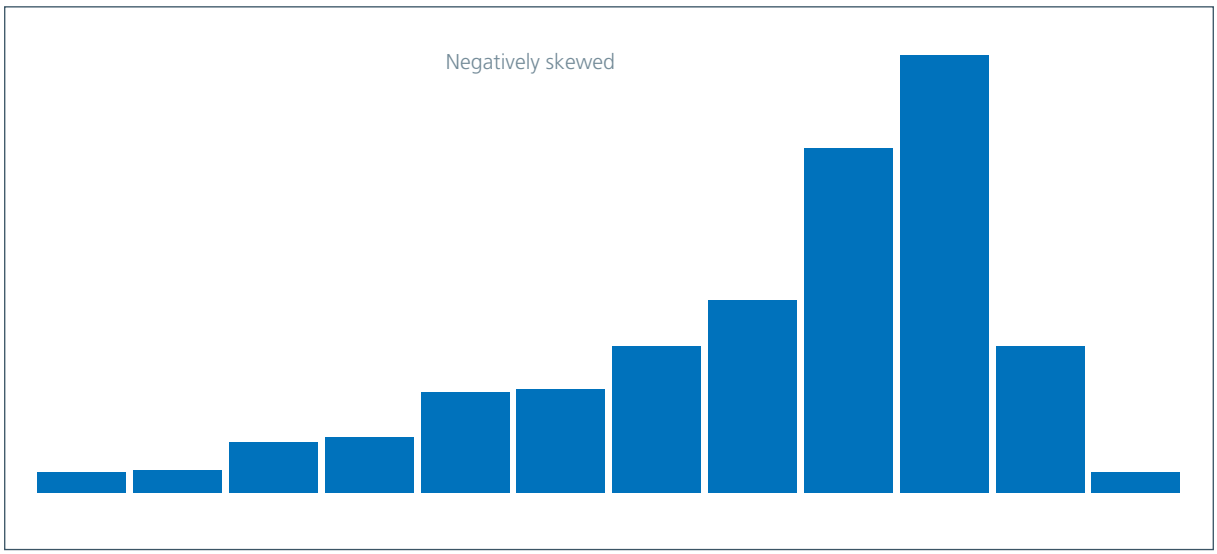
One of the key characteristics of normal distribution is the relationship between the shape of the curve and the standard deviation (plus or minus three standard deviations from the mean).

Evaluating the distribution for normality affects which statistical tests you can perform. If the data is not normally distributed, you cannot calculate control limits and create a **statistical process control (SPC)** chart to identify special causes of **variation** in a process under review. Neither can you highlight improvement opportunities.

Data may be skewed to the left or to the right. If the histogram shows a long tail of data on the left hand side, this is called left or negatively skewed distribution (histogram 2). If the tail appears on the right hand side, this is called right or positively skewed distribution.

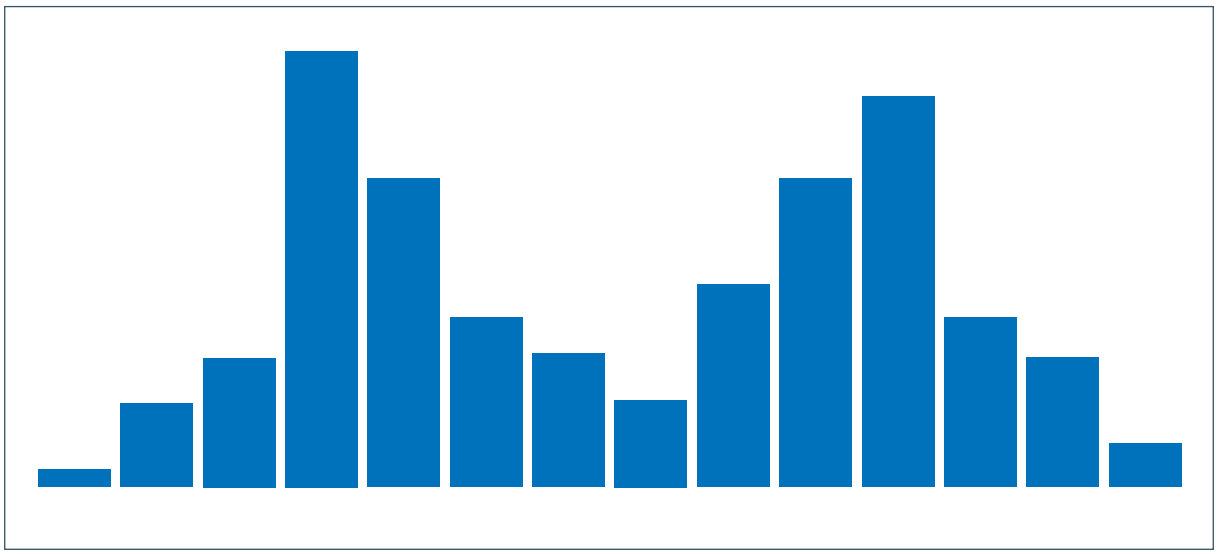
Seriously skewed data may be an indication that there are inconsistencies in the process. For instance, measures may tend to artificially 'pile up' just on the good side of some target figure.

**Figure 4: Histogram 2**




If the shape of your histogram shows twin or multiple peaks (histogram 3), this is an indication that the data is coming from two or more different sources, eg shifts, departments, wards, pathways, tests. In this case you may need to separate the data out so that one histogram represents one set of data.

**Figure 5: Histogram 3**



## What next?

Evaluate the distribution of the set of data to inform your improvement effort. If there is normal distribution, apply certain statistical tests such as **statistical process control (SPC)** to identify **variation** in the process and special or common causes. In the case of a twin peak distribution, re-work and separate the data.

## Additional resources

Brue, G (2002) *Six Sigma for Managers*, McGraw Hill

George, M *et al* (2005) *The Lean Six Sigma Pocket Toolbook*, McGraw Hill

Smith, H, Burns, G, Linsky, J and Pledger, K (Ed.) (2011) *Revise Edexcel GCSE Mathematics – Specification A Linear – Revision Guide (Higher)*, Pearson Schools

## Background

Descriptive statistics and data display are used to enable you to analyse the data under review and gain an understanding of the problems. A histogram provides basic information about the distribution and properties of a set of data. It shows you the types of frequency plots, which allow you to interpret the patterns they reveal.

What is important when evaluating the normality recommended for any set of continuous data is to learn about the basic characteristics and to evaluate whether you can apply certain statistical tests.