Six Sigma – Part 3: DMAIC: Analyzing

An Online Continuing Education Course for Engineers

Course Number: P-6001

Credit: 6 Hours / 6 PDH / 6 CPD

Module 1

Depicting and Analyzing Data through Charts and Graphs

In This Module

- Understanding the power inherent in basic charts and graphs
- Creating and analyzing variation and distributions through histograms or dot plots
- Comparing distributions with box and whisker plots
- Exploring variable relationships with scatter plots
- Using process behavior charts to see true performance

Data are the medium in which you, as a Six Sigma practitioner, work. Therefore, how you depict and analyze your data is critical.

The most important — and often the most powerful — tools for analyzing and communicating data are graphical; they display the nitty-gritty of what's going on in the company through pictures, charts, and graphs. These basic graphical tools, called *graphs* or *plots,* are much simpler to use than their strictly numeric counterparts (such as a table of raw data). Pictures of data often serve as a more intuitive way of gaining the same insights that you would from crunching numbers, and pictures make communicating your gained insight to others easy.

The chief purpose of plotting and charting data is to graphically show the central tendency and the spread of variation in a measured item of interest. You can accomplish this task in a couple of different ways; each has its advantages and disadvantages. In this module, we show you how to visually display your data through histograms or dot plots, box and whisker plots, scatter plots, and process behavior charts. You also figure out how to pick up on the signs and clues that the data give you when they're portrayed graphically.

Using visual materials to communicate data is your best way of getting improvement team members to be integral parts of the Six Sigma breakthrough process. When team members can see the reality of performance for themselves, they're more motivated to contribute and participate in measurement and improvement efforts. Also, your visual pictures are an effective and essential prop for communicating your project details to management.

Checking Out Dot Plots and Histograms

Both dot plots and histograms give you lots of information about the variation of a critical characteristic in a process. A *dot plot* shows the scatter and grouping of a data from a single characteristic using (no surprise here) dots. A *histogram* takes the data from the dot plot and replaces the dots with bars. The following sections show you how to generate these helpful graphics and understand what the graphs are telling you.

Creating your own dot plots and histograms

After collecting measurements or data for a characteristic, create a *dot plot* for it by using the following steps:

1. **Create a horizontal line that represents the scale of measure for the characteristic.**

   This scale should be in whatever measure best quantifies the aspect of the characteristic you're interested in — for example, millimeters for length, pounds for weight, minutes for time, or number of defects found on an inspected part.

2. **Divide the horizontal scale of measure into equal chunks or "buckets" along its length.**

   Select a bucket width that creates about 10 to 20 equal divisions between the largest and smallest observed values for the characteristic.

3. **For each observed measurement of the characteristic, locate its value along the horizontal scale and place a dot for it in its corresponding bucket.**

   If another observed measurement falls into the same "bucket," stack the second (or third, or fourth) dot above the previous one.

4. **Repeat Step 3 until you've placed all the observed measurements onto the plot.**

To create a histogram (so that you can impress your peers with a graph that has a much more complicated-sounding name), replace each of the stacks of dots with a solid vertical bar of the same height as its corresponding stack of dots. ***Note:*** The vertical dimension on a dot plot or histogram is sometimes called *frequency* or *count* (refer to the following section). Because it's pretty straightforward, the label of the vertical scale is often left off the plot, which is why we don't include it in the process here.

Interpreting dot plots and histograms

A dot plot and its fancy cousin, the histogram, offer ready access to a wealth of information about the variation of a characteristic's performance. The following points are a few aspects of a dot plot or histogram to note:

- **Plot height**: The frequency — the height of dots or the bar — in a dot plot or histogram indicates how often the corresponding value on the horizontal axis was observed.
- **Variation shape:** The shape of the variation on a histogram comes in three basic flavors: normal, uniform, and skewed. Figure 1-1 shows a variation shape that is *normally* distributed, or bell-shaped. For a normal distribution, most of the observed values of the characteristic are close to a central point, with fewer and fewer values appearing as you get farther away from the central tendency.
  Figure 1-2 shows a uniformly distributed variation for a characteristic. With *uniform* distribution, the variation is evenly spread out across a bounded range. That is, you're just as likely to observe a value for a characteristic at one end of the interval as you are at the other, or anywhere in between. Figure 1-3 shows a skewed distribution shape. A *skewed* distribution is a variation shape that isn't symmetrical; one side of the distribution extends out farther than the other side.

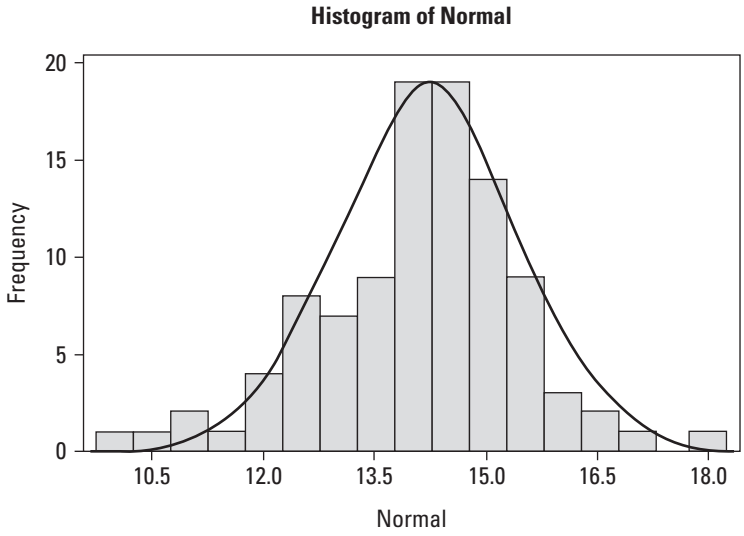

Figure 1-1: Histogram showing normally distributed variation.

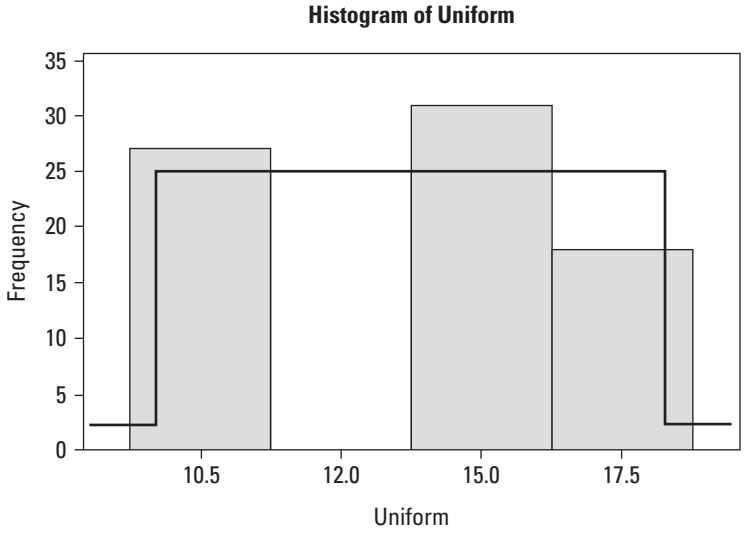

Figure 1-2: Histogram showing a uniformly distributed variation of a characteristic.

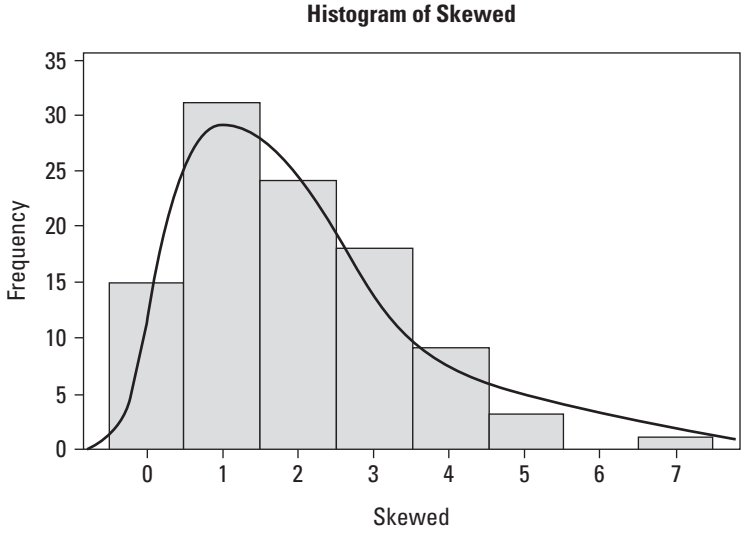

Figure 1-3: Histogram showing a skewed distribution.

- **Variation mode:** The *mode* of a distribution is its most often repeated value, or in other words, its peak. Usually, the variation in a characteristic has a single peak, as seen in Figure 1-4.
  But sometimes, a characteristic displays two or more modes, as shown in Figure 1-5, because two or more values dominate the variation. A histogram showing two or more distinct peaks is *multi-modal.* Multiple

major peaks aren't usual; this situation typically means that a factor affecting the characteristic's performance is causing the entire system to behave schizophrenically.

When you encounter a multi-modal distribution, always dig deeper to discover what factor or factors are causing the characteristic's schizophrenic behavior.

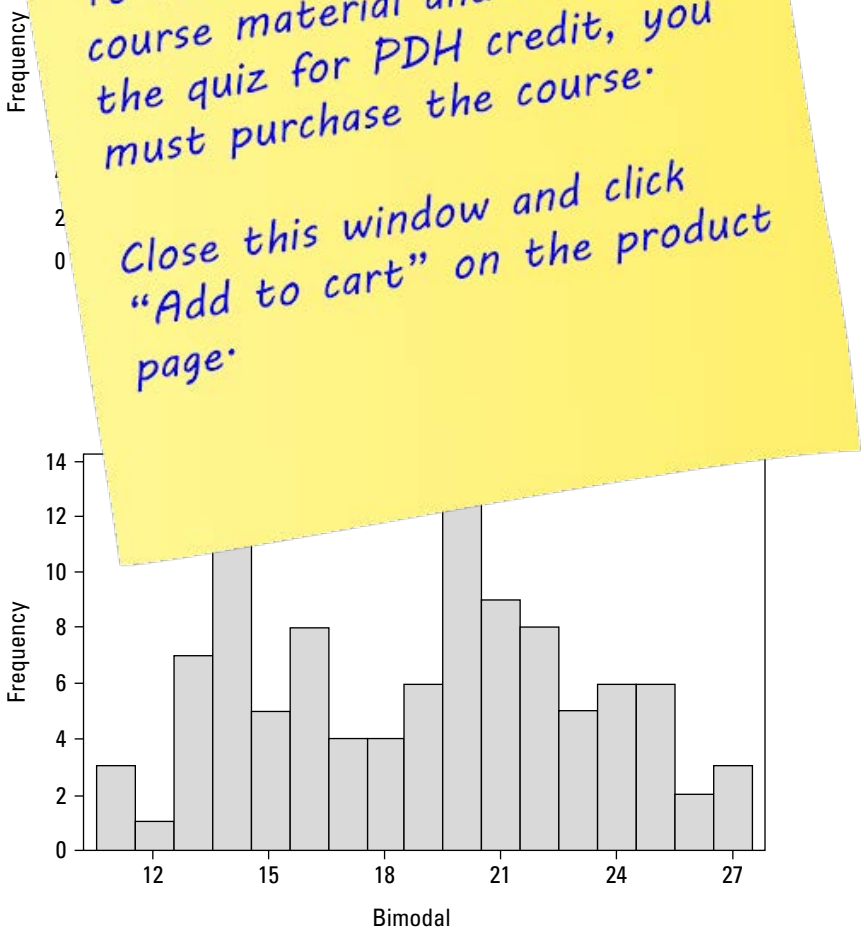

Figure 1-4: Histogram showing a single variation mode.

Figure 1-5: Histogram showing a multi-modal distribution.