Notes:

Histograms

Key Learning Points

1. Describe the importance of using Histograms.
2. Explain how to create and interpret Histograms.
3. Utilize Histograms in improvement projects.

What is a Histogram?

A histogram is a chart that uses bars to display variation in a single characteristic. Patterns in the variation often reveal new facts about the process.

Values

Values in a set of data almost always show variation. Variation is everywhere. It is inevitable in the output of any process. It is impossible to keep all factors in a constant state all the time.

Variation

Variation displays a pattern. Different phenomena will have different variation, but there is always some pattern to the variation.

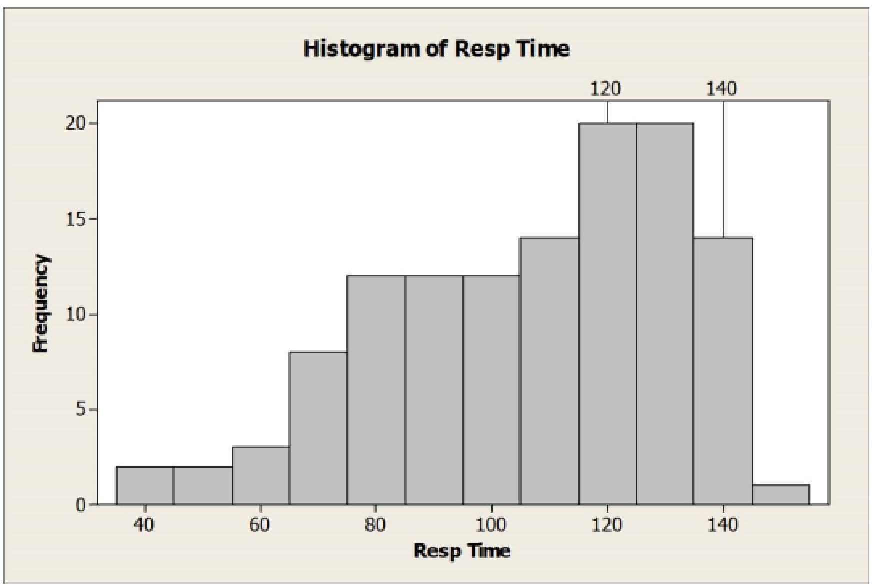
Patterns

These patterns of variation in data are called "distributions." For this purpose, you simply want to point out that there are usually discernible patterns in the variation, and these patterns often tell a great deal about the cause of a problem. Identifying and interpreting these patterns is key to working with histograms.

Histogram Data

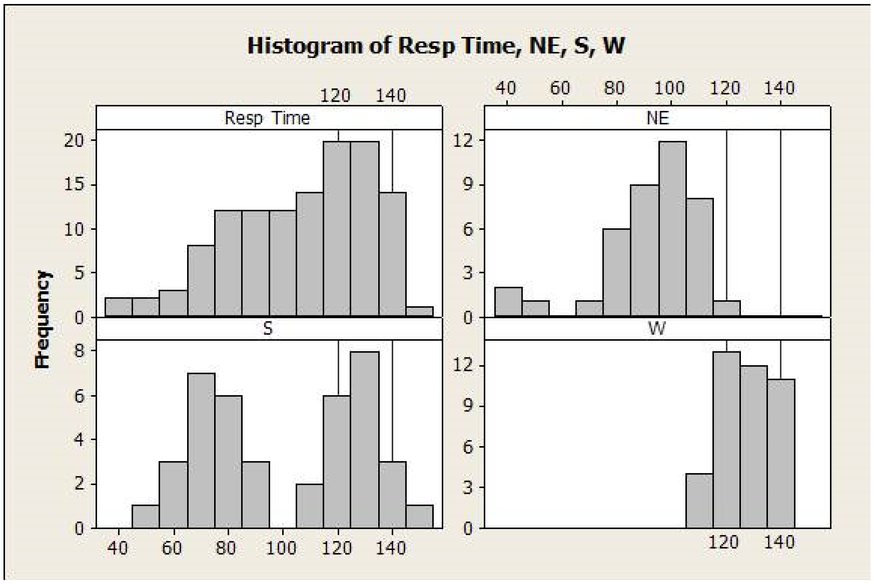
Region	Resp Time	NE	S	W
NE	80	80	124	140
NE	95	95	122	144
NE	115	115	117	122
NE	110	110	144	120
NE	93	93	118	116
NE	105	105	119	110
NE	42	42	128	138
NE	78	78	120	136
NE	111	111	134	138
NE	100	100	126	135
NE	105	105	144	140
NE	110	110	139	128
NE	107	107	128	126
NE	100	100	130	130
NE	95	95	108	122
NE	80	80	128	116
NE	35	35	110	106
NE	102	102	133	128

Histogram



Notes:

Stratified Histogram



Characteristics

Patterns of variation are difficult to see in simple tables of numbers, histograms show these patterns. Three important aspects of histograms are:

- Its center
- Its width
- Its shape

Steps in Constructing Histograms

1. Determine the high value, low value, and range.
   a. Range = high value – low value
2. Decide on the number of cells or bars.
3. Calculate the approximate cell width.
   a. Approximate Cell Width = range/number of cells
4. Round the cell width to a convenient number.
5. Construct the cells by listing the cell boundaries.
6. Tally the number of data points in each cell.
7. Draw and label the horizontal axis.
8. Draw and label the vertical axis.
9. Draw the bars to represent the number of data points in each cell.
10. Title the chart, indicate the total number of data points, and show nominal

Notes:

values and limits (if applicable).

11. Identify and classify the pattern of variation.
12. Develop an explanation for the pattern.

Interpretation

Concept 1

Values in a set of data almost always show variation. Will the measurement be a constant, or will there be some variation in the data?

Examples

- The number of pieces of candy in a one-pound bag
- The exact concentration of an intravenous medication
- The time required to repair an appliance for a customer
- The number of minutes required to process an invoice

In each case, the measurement will show some variation; few values will be exactly the same

Concept 2

Variation displays a pattern. Different phenomena will have different variation, but there is always some pattern to the variation.

These patterns of variation in data are called "distributions." For your purposes, you simply want to point out that there are usually discernible patterns in the variation, and these patterns often tell a great deal about the cause of a problem.

There are three important characteristics of a histogram:

- Its center
- Its width
- Its shape

Concept 3

Patterns of variation are difficult to see in simple tables of numbers. While there is a pattern in the data, it is difficult for eyes and minds to see it. It is easy to conclude erroneously.

Concept 4

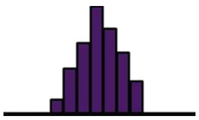
Patterns of variation are easier to see when the data are summarized pictorially in a histogram.

Notes:

Interpreting Patterns of Distribution

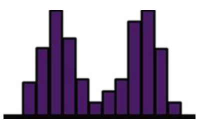
Bell Shaped

A symmetrical shape with a peak in the middle of the range of the data. This is the normal, natural distribution of data from a process. Deviations from this bell shape may indicate the presence of complicating factors or outside influences. While deviations from a bell shape should be investigated, such deviations are not necessarily bad. As you will see, some non-bell distributions are to be expected in certain cases.



Double-Peaked

A distinct valley in the middle of the range of the data with peaks on either side. This pattern is usually a combination of two bell-shaped distributions and suggests that two distinct processes are at work.



Plateau

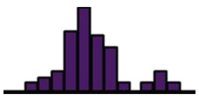
A flat top with no distinct peak, and slight tails on either side. This pattern is likely to be the result of many different bell-shaped distributions with centers spread evenly throughout the range of the data. Diagram the flow and observe the operation to identify the many different processes that are at work. An extreme case occurs in organizations that have no defined processes or training—everyone does the job his or her own way. The wide variability in process leads to the wide variability observed in the data. Defining and implementing standard procedures will reduce this variability.

Notes:

Notes:

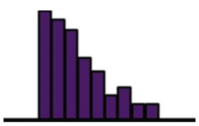
Isolated-Peaked

A small, separate group of data in addition to the larger distribution. Like the double-peaked distribution, this pattern is a combination and suggests that two distinct processes are at work. The small size of the second peak indicates an abnormality, something that doesn't happen often or regularly.
Look closely at the conditions surrounding the data in the small peak to see if you can isolate a particular time, machine, input source, procedure, operator, etc. Such small isolated peaks in conjunction with a truncated distribution may result from the lack of complete effectiveness in screening out defective items. It is also possible that the small peak represents errors in measurements or in transcribing the data. Recheck your measurements and calculations.



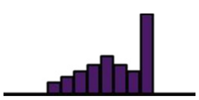
Truncated

An asymmetrical shape in which the peak is at or near the edge of the range of the data, and the distribution ends very abruptly on one side and tails off gently on the other. The illustration shows truncation on the left side with a positively skewed tail. Of course, you may also encounter truncation on the right side with a negatively skewed tail. Truncated distributions are often smooth, bell-shaped distributions with a part of the distribution removed, or truncated, by some external force such as screening, 100 percent inspection, or a review process. Note that these truncation efforts are an added cost and are, therefore, good candidates for removal.



Edge-Peaked

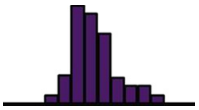
A large peak is appended to an otherwise smooth distribution. This shape occurs when the extended tail of the smooth distribution has been cut off and lumped into a single category at the edge of the range of the data. This shape very frequently indicates inaccurate recording of the data, e.g., values outside the "acceptable" range are reported as being just inside the range.



Skewed

An asymmetrical shape in which the peak is off-center in the range of data and the distribution tails off sharply on one side and gently on the other.
The skewed pattern typically occurs when a practical limit, or a specification limit, exists on one side and is relatively close to the target value. In these cases, there simply are not as many values available on one side as there are on the other side.

Practical limits occur frequently when the data consist of time measurements or counts of things.
One-sided specification limits (a maximum or minimum value only) also frequently give rise to skewed distributions. Such skewed distributions are not inherently bad, but a team should question the impact of the values in the long tail. Could they cause customer dissatisfaction, e.g., long waiting times? Could they lead to higher costs, e.g., overstaffing, excess overtime? Could the extreme values cause problems in downstream operations? If the long tail has a negative impact on quality, the team should investigate and determine the causes for those values.



Comb

High and low values alternating in a regular fashion. This pattern typically indicates measurement error, errors in the way the data were grouped to construct the histogram, or a systematic bias in the way the data were rounded off. This might also be a type of plateau distribution, but the regularity of alternating highs and lows is a warning of possible errors in data collection or in histogram construction. Review the data-collection procedures and the construction of the histogram before considering possible process characteristics that might cause the pattern.

When Should A Histogram Be Used?

The histogram is a useful tool when a team is faced with the task of analyzing data that contain variation. We know
intuitively that the variation will usually follow some pattern, but the pattern is often hard to see from the table of numbers. Because it is a “picture” of the data, a histogram enables us to see this pattern of variation.

Pitfalls to Avoid

- Teams may fail to use histograms if they use the following assumptions: Team members already know the cause.
- Simple numerical index, such as the average or range, provide an adequate summary of the data, i.e., consider the two sets of data:
    - 0 0 0 0 0 10 10 10 10 10
    - 5 5 5 5 5 5 5 5 5

Histograms are not sensitive enough to allow you to make firm conclusionsabout small differences in variability or the locations of peaks in distributions.

Histograms can also obscure the time differences among data sets. In quality control, a histogram of a process run tells only one part of a long story. There is a need to keep reviewing the histograms and control charts for consecutive process

Notes:

runs over an extended time to gain useful knowledge about a process.

Notes: