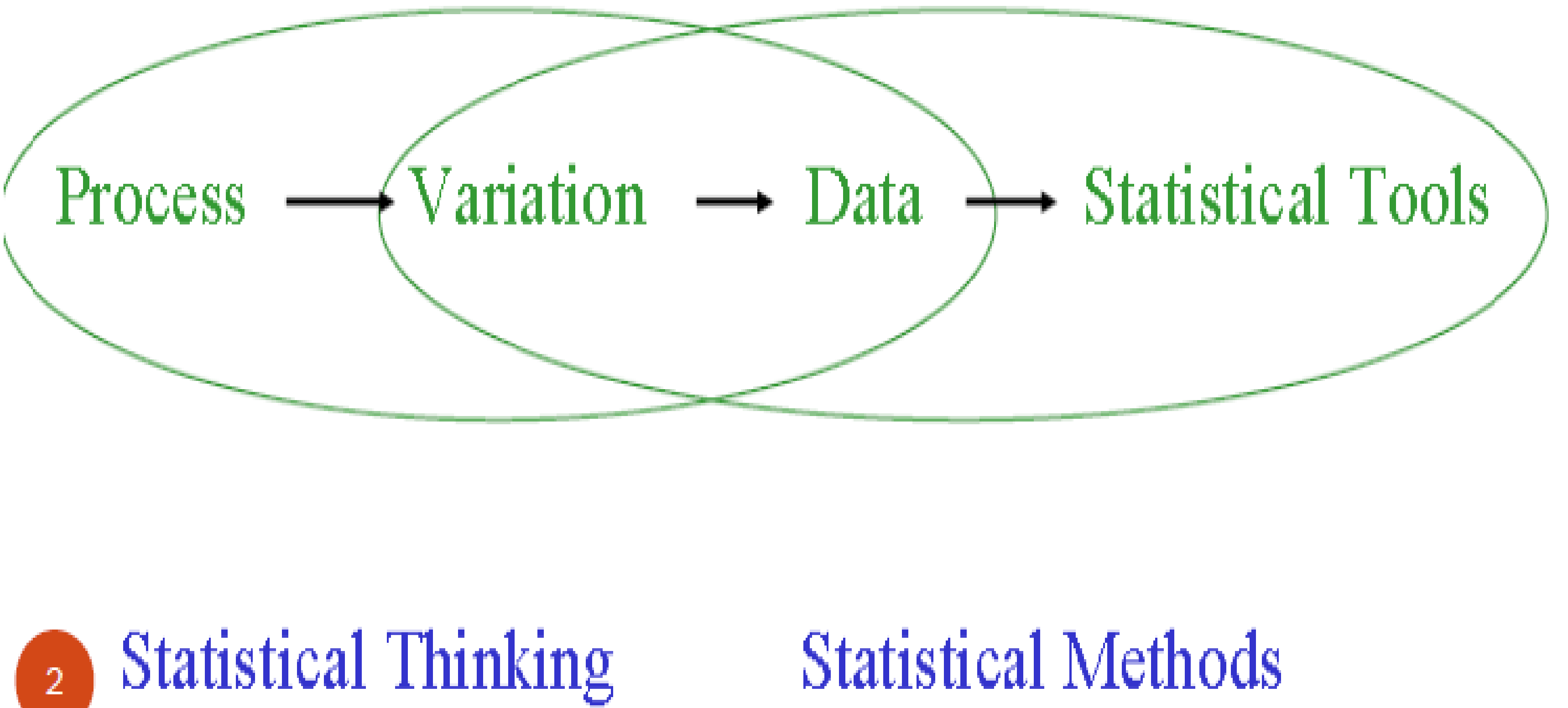
CHAPTER TWO:
STATISTICAL TOOLS FOR PROBLEM SOLVING

2.1. Statistical thinking

Statistical Thinking is a philosophy of learning and acting based on the following fundamental principles:

- All work occurs in a system of interconnected processes,
- Variation exists in all processes, and
- Understanding and reducing variation are keys to success.

Statistical Thinking and Methods



Key concepts of statistical thinking

- Process and Systems thinking
- Variation
- Analysis to increase knowledge
- Taking action
- Improvement

Role of data in statistical thinking

- Quantify variation
- Measure effects

Statistical thinking without a process view

- People have problems understanding the problem and their role in its solution
- It is difficult to define the scope of the problem
- It is difficult to get root causes
- People get blamed when the process is the problem
- Process management is ineffective
- Improvement is slowed

Statistical thinking without understanding variation

- There will be management by the last data point
- There's lots of fire fighting
  - ✓ Using special cause methods to solve common cause problems
- Understanding the process is handicapped
  - ✓ Learning is slowed
- Process management is ineffective
- Improvement is slowed

Statistical thinking without data

- Everyone is an expert
- Historical memory is poor

- Difficult to get agreement on:
  - ✓ What the problem is
  - ✓ What success looks like
  - ✓ Progress made
- Process management is ineffective
- Improvement is slowed

2.2 Measurement and Data Collection

If we adopt the definition of quality as ‘meeting the customer requirements’, there is already the need to consider the quality of design and the quality of conformance to design.

To achieve quality therefore requires:

- An appropriate design;
- Suitable resources and facilities (equipment, premises, cash, etc.);
- The correct materials;
- People, with their skills, knowledge and training;

- An appropriate process;
- Sets of instructions;
- Measures for feedback and control.
- All are data which need to be measured and collected

Types of data

1. Categorical data – Data for discreet variables
2. Continuous data – Data for continuous variables

1. Categorical variables:

Numbers and proportions in each category – 5 points Likert scale – Highly satisfied, Satisfied, Neither satisfied nor dissatisfied, dissatisfied, highly dissatisfied)

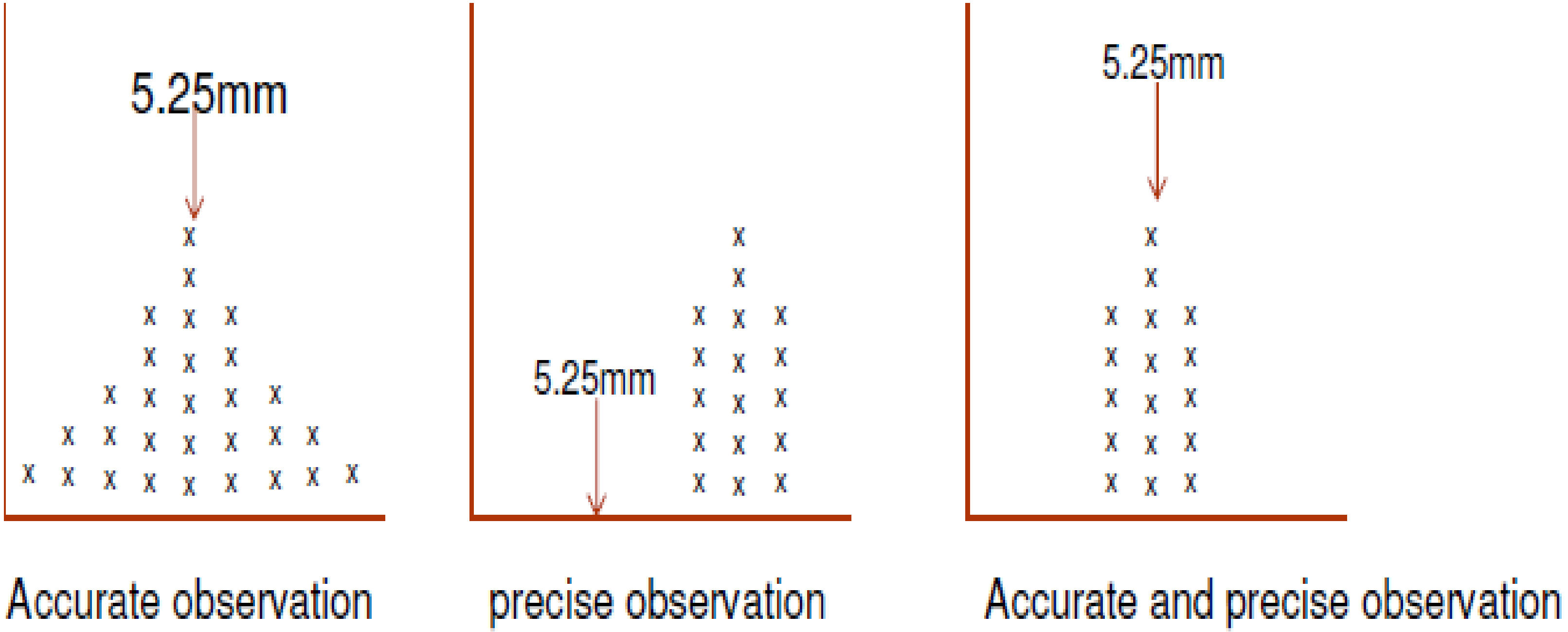
2. Continuous variables

1. Distributions
    - Normal (Gaussian)
    - Non-parametric
2. Central tendency
    - Mean
    - Median
3. Scatter
    - Standard deviation
    - Range
    - Inter-quartile range
    - Standard error of the mean

Accuracy and Precision

- The precision of a data set or a measuring instrument refers to the degree of variability of the observations. Observations may be off the target value but still considered as precise.
- The accuracy of a data set or a measuring instrument refers to the degree of uniformity of the observations around a desired value such that, on average, the target value is realized.

Example: Target value = 5.25mm



2.3. Populations and Samples – Sampling Issues

Sampling is the process of obtaining samples.

A sampling design is a description of the procedure by which the observations in a sample are to be chosen. It does not necessarily deal with the measuring instrument to be used.

Example- every tenth item produced

Population – the entire group of elements from which data will be collected.

Element is an object (or group of objects) for which data or information is gathered.

A Sampling Unit is an individual element or a collection of non-overlapping elements from the population

A Sampling frame – is a list of all sampling units.

Sample Space – a set that consists of all possible outcomes of a random experiment.

Sampling Design and Schemes

A major objective of any sampling design or scheme is to select the sample in such a way as to accurately portray the population from which it is drawn. After all, a sample is supposed to be representative of the population.

- Sampling in general has certain advantages, if the measurement requires destroying the item being measured (destructive testing).

If we draw an object from urn, we have the choice of replacing or not replacing the object into urn before we draw again. In the first case a particular object can come up again and again, whereas in the second it can come up only once.

- **Sampling with replacement** - Sampling where each member of a population may be chosen more than once.

- **Sampling without replacement** – Sampling where each member cannot be chosen more than once.

- A finite population that is sampled with replacement can theoretically be considered infinite since samples of any size can be drawn without exhausting the population. For most practical purposes, sampling from a finite population that is very large can be considered as sampling from as infinite population.

2.4. Fundamental Statistical Measures

Measures of Central Tendency

- A measure of central tendency gives a single value that acts as a representative or average of the values of all the outcomes of your experiment.
- The main measure of central tendency we will use is the arithmetic mean. While the mean is used the most, two other measures of central tendency are also employed. These are the median and the mode. Each has advantages and disadvantages, depending on the data and the intended purpose.

- **Mean** – if we are given a set of n numbers of values, the mean is the average of these n values.
- **Median** – is that value x for which $P(X<x)<=1/2$ and $P(X>x)<=1/2$. In other words, the median is the value where half of the values of the total sample are larger than the median, and half of them are smaller than the median.

Example: Consider the following set of integers:

$S = \{1, 6, 3, 8, 2, 4, 9\}$

- If we want to find the median, we need to find the value, x, where half the values are above x and half the values are below x. begin by ordering the list in ascending order.

S = {1, 2, 3, 4, 6, 8, 9}

4 is at the middle way of the list and it is the median.

- If the number of values are even number, we will find two values at the middle way of the list and the median can be any number between these two numbers but, commonly it is taken the average value of these numbers:

Example: S = {1, 2,3, 4, 6, 8, 9, 12}

4 and 6 are at the middle way of the list and the median is 5 (the average)

Mode – is the vale that occurs most often.

Example: consider the following rolls of a ten-sided die:

R = {2, 8, 1, 9, 5, 2, 7, 2, 7, 9, 4, 7, 1, 5, 2}

The number that appears the most is 2. therefore, the mode is 2.

If there are more than one values which appear the most:, the distribution is called

- Bimodal – if there are two values
- Tri-modal – if there are three values
- Multimodal – if there are more than three values

Measures of Variation (Dispersion)

Consider the following two sets of integers:

S={5, 5, 5, 5, 5,5} and R={0,0,0, 10, 10, 10}

The mean is 5 for both S and R. but they are two vastly different data sets. We need another descriptive statistic besides a measure of central tendency, which we shall call a measure of variation or measure of dispersion.

We can measure the dispersion or scatter of the values of our data set about the mean of the data set.

- If the values tend to be concentrated near the mean, then this measure shall be small, while if the values of the data set tend to be distributed far from the mean, then the measure will be large.
- The three measures of variations that are usually used are called the range, variance and standard deviation.

Variance – is denoted by σ^2

- For a set of n numbers $x_1, x_2, \ldots, x_n$ and if μ is the mean of the values, the variance is given by:

$$\sigma^2 = \frac{[(X_1 - \mu)^2 + (X_2 - \mu)^2 + \ldots + (X_n - \mu)^2]}{n}$$

The variance is a nonnegative number.

Standard deviation -The positive square root of the variance.

Example: T= {75, 80, 82, 87, 96}

$\mu=84$, $\sigma^2 = 50.8$ and standard deviation = 7.1274118

Range – the difference between the maximum and minimum values of a measured parameter.

2.5. Quality Control Tools for Data Collection and Analysis

- The purpose of statistical quality control is to ensure, in a cost efficient manner, that the product shipped to customers meets their specifications.
- Statistical Process Control (SPC) is an analytical decision making tool which allows you to see when a process is working correctly and when it is not.

There are many ways to implement process control. Key monitoring and investigating tools include:

1. Histograms,
2. Check sheets,
3. Pareto charts,
4. Cause and effect diagrams,
5. Scatter diagrams, and
6. Control charts.
7. Flowchart.

Histogram

- The histogram is a bar chart showing a distribution of variables.
- This tool helps identify the cause of problems in a process by shape of the distribution as well as the width of the distribution.
- The histogram clearly portrays information on location, spread, and shape regarding the functioning of the physical process.
- It can also help suggest both the nature of, and possible improvements for, the physical mechanisms at work in the process.

Creating a frequency distribution table

Step 1: Identify an interval that is wide enough to contain all the data.

Step 2: Subdivide the interval identified in Step 1 into class intervals of equal width. The class intervals will serve as the categories.

Step 3: Set up a table with three columns for the following: class interval, tally, and frequency. (The tally column can be removed in the final table.)

Step 4: To complete the table, determine the frequency with which data values fall into each class interval.

Example:

below are thickness measurements, in millimeters, from a sample of 25 plastic sheet used in the manufacture of device cover. Notice that it is difficult to extract much information from staring at these numbers. So forming histogram is important.

0.402	0.496	0.533	0.387	0.384
0.528	0.411	0.367	0.462	0.499
0.539	0.546	0.425	0.457	0.586
0.558	0.588	0.425	0.437	0.479
0.427	0.485	0.443	0.441	0.658

Solution

Thickness (mm)	Tally	Frequency
0.30 – 0.35		0
0.35 – 0.40	|||	3
0.40 – 0.45	~~||||~~ |||	8
0.45 – 0.50	~~||||~~ |	6
0.50 – 0.55	||||	4
0.55 – 0.60	|||	3
0.60 – 0.65		0
0.65 – 0.70	|	1

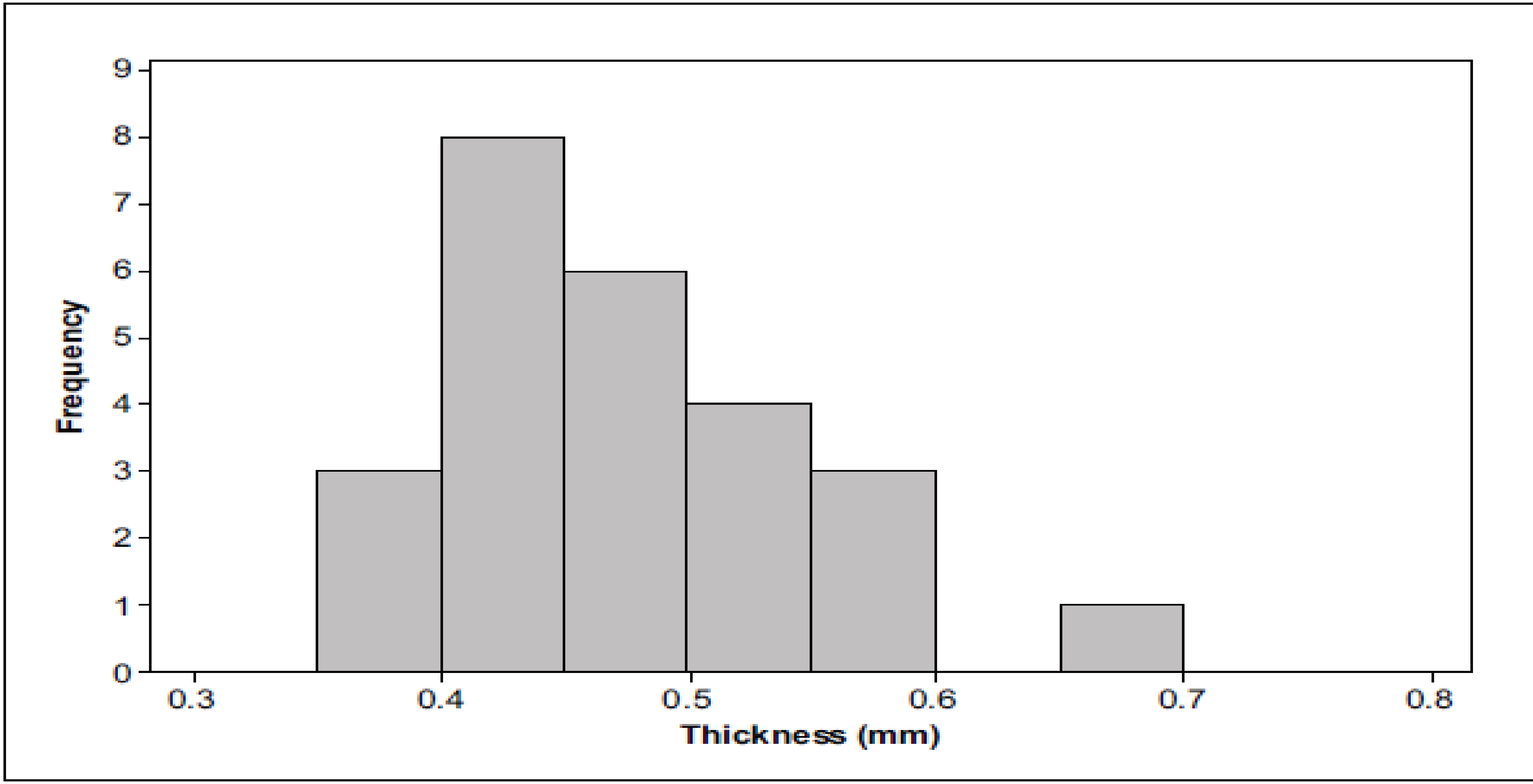
A completed frequency distribution table

Creating a histogram from a frequency distribution

Step 1: Draw a set of axes. On the horizontal axis, mark the boundaries of the class intervals. On the vertical axis, set up a scale appropriate for the frequencies. (Later this scale can be changed to proportion or percent.)

Step 2: Label the horizontal axis with the name of the variable being measured and the units.

Step 3: Over each interval, draw a rectangle with the interval as its base. The height of the rectangle should match the frequency of data contained in that interval.



Histogram representing frequency distribution

2. Check Sheet

- A check sheet is one of the seven basic quality tools. Data collection can often become an unstructured and messy exercise.
- It is a simple form you can use to collect data in an organized manner and easily convert it into readily useful information.
- Data collection is important because it is the starting point for statistical analysis.

Characteristics of check sheet

- Data can be recorded easily
- Data can be understood easily
- Can prevent missing data
- Can determine source of problem
- Can facilitate quick resolving
- Can check several items at once
- Can facilitate classification (stratification) of
- data.

<table>
<tr><td colspan="2">Product No.:</td><td colspan="4">Product Name:</td></tr>
<tr><td colspan="2">Date:</td><td>Shift:</td><td colspan="3">Operators Name:</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td rowspan="2">Problems/Causes of problems</td><td colspan="5">Days/Weeks/Months</td></tr>
<tr><td>1</td><td>2</td><td>…</td><td>m</td><td>Total</td></tr>
<tr><td>Problem 1/ Cause 1</td><td>No. of Occurrences</td><td>| | | |</td><td>…</td><td>| |</td><td></td></tr>
<tr><td>Problem 2/ Cause 2</td><td>| | |</td><td>| | | | |</td><td></td><td></td><td></td></tr>
<tr><td>…</td><td>| | | | | |</td><td>|</td><td></td><td></td><td></td></tr>
<tr><td>Problem n/ Cause n</td><td>| |</td><td>| | |</td><td></td><td>Occurrence</td><td></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="5">Total number Inspected</td><td></td></tr>
<tr><td colspan="5">Total number rejected</td><td></td></tr>
<tr><td colspan="5">Rejection rate (%)</td><td></td></tr>
</table>

Figure: Structure of Check Sheets

- After recording the data like this, the team will discuss for each occurrences and their causes in detail. After thorough discussion on the data, solutions for the problems and how to sustain improvements are proposed.

Phases of Using Check Sheet

It has two phases, namely:

1. Define – specify from which section of the company to collect data, what problems to address, and the time period.
2. Measure – record the occurrences of each problem in each specific period.

Guidelines to us check sheet

- Use the Check Sheet when data can be observed and collected repeatedly by the same person or at the same location.
- Use the Check Sheet when collecting data on the frequency or patterns of events, problems, defects, defect location, defect causes, etc.

3. Pareto Analysis

Helps prioritize problems by arranging them in decreasing order of importance. In an environment of limited resource, these diagrams help companies decide on the order in which they should address problems.

- It is a bar graph used to arrange information in such a way that priorities for process improvement can be established.

Steps to construct a Pareto diagram:

Step 1:

Determine the categories and the units for comparison of the data, such as frequency, cost, or time.

Step 2:

Total the raw data in each category, then determine the grand total by adding the totals

of each category.

Step 3: Re-order the categories from largest to smallest.

Step 4:

Determine the cumulative percent of each category (i.e., the sum of each category plus all categories that precede it in the rank order, divided by the grand total and multiplied by 100).

Step 5:

Draw and label the left-hand vertical axis with the unit of comparison, such as frequency, cost or time.

Step 6:

Draw and label the horizontal axis with the categories. List from left to right in rank order.

Step 7:

Draw and label the right-hand vertical axis from 0 to 100 percent. The 100 percent should line up with the grand total on the left hand vertical axis.

Step 8:

Beginning with the largest category, draw in bars for each category representing the total for that category.

Step 9:

Draw a line graph beginning at the right hand corner of the first bar to represent the cumulative percent for each category as measured on the right-hand axis.

Step 10: Write any necessary items on the diagram.

Step 11: Analyze the chart. Usually the top 20% of the categories will comprise roughly 80% of the cumulative total.

Example

The following table shows the different types of defect and the total number of items that are occurred on selected products in an ideal company ABC. Use the Pareto analysis to determine the vital few cause, which results the majority of the problem.

Table Number of defects observed

Type of Defect	Number of Defects
Crack	10
Scratch	42
Stain	6
Strain	104
Gap	4
Pinhole	20
Others	14
Total	200

Solution:

Step 1:

1. Decide what problems are to be investigated and how to collect the data.

2. Decide what kind of problems you want to investigate.

Example: Defective items, losses in monetary terms, accidents occurring.

3.Decide what data will be necessary and how to classify them. *Example:* By type of defect, location, process, machine, worker, method.

Note: Summarize items appearing infrequently under the heading "others."

4.Determine the method of collecting the data and the period during which it is to be collected.

Note: Use of an investigation form is recommended.

Step 2:

Design a data tally sheet listing the items, with space to record their totals.

Step 3:

Make a Pareto diagram data sheet listing the items, their individual totals, cumulative totals, percentages of overall total, and cumulative percentages.

Step 4:

Arrange the items in the order of quantity, and fill out the data sheet.

***Note**:* The item "others" should be placed in the last line, no matter how large it is. This is because it is composed of a group of items each of which is smaller than the smallest item listed individually.

Table : Data Sheet for Pareto Diagram

Type of Defects	Number of Defects	Cumulative Total	Percentage of overall Total	Cumulative Percentage
Strain	104	104	52	52
Scratch	42	146	21	73
Pinhole	20	166	10	83
Crack	10	176	5	88
Stain	6	182	3	91
Gap	4	186	2	93
Others	14	200	7	100
Total	200	-	100	-

Step 5:

Draw Left-hand vertical axis and mark this axis with a scale from 0 to the overall total two vertical axes and a horizontal axis.

Step 6:

Draw horizontal axis, and divide this axis into the number of intervals to the number of items classified.

Step7: Draw Right-hand vertical axis and mark this axis with a scale from 0 % to 100 %.

Step 8: Construct a bar diagram.

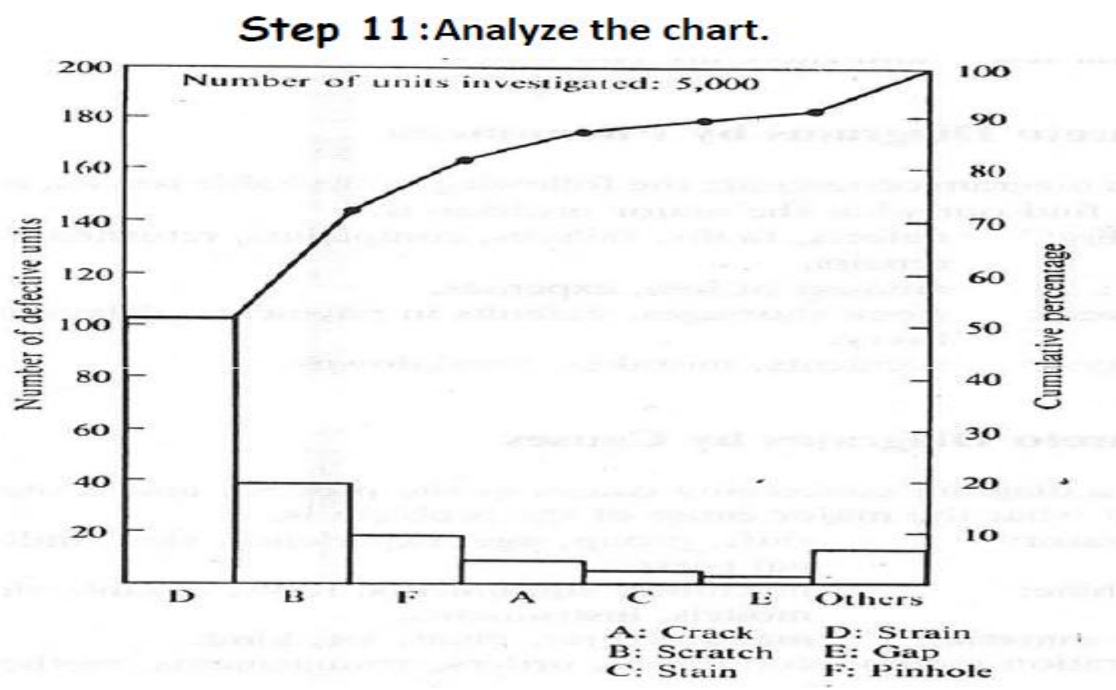
Step 9: Draw the cumulative curve (Pareto curve) as shown in figure 6.2. Mark the cumulative values (cumulative total or cumulative percentage), above the right-hand intervals of each item, and connect the points by a solid line.

Step 10: Write any necessary items on the diagram.

1.Items concerning the diagram as title, significant quantities, units, name of drawer

2.Items concerning the data as period, subject and place of investigations, total number of data etc.

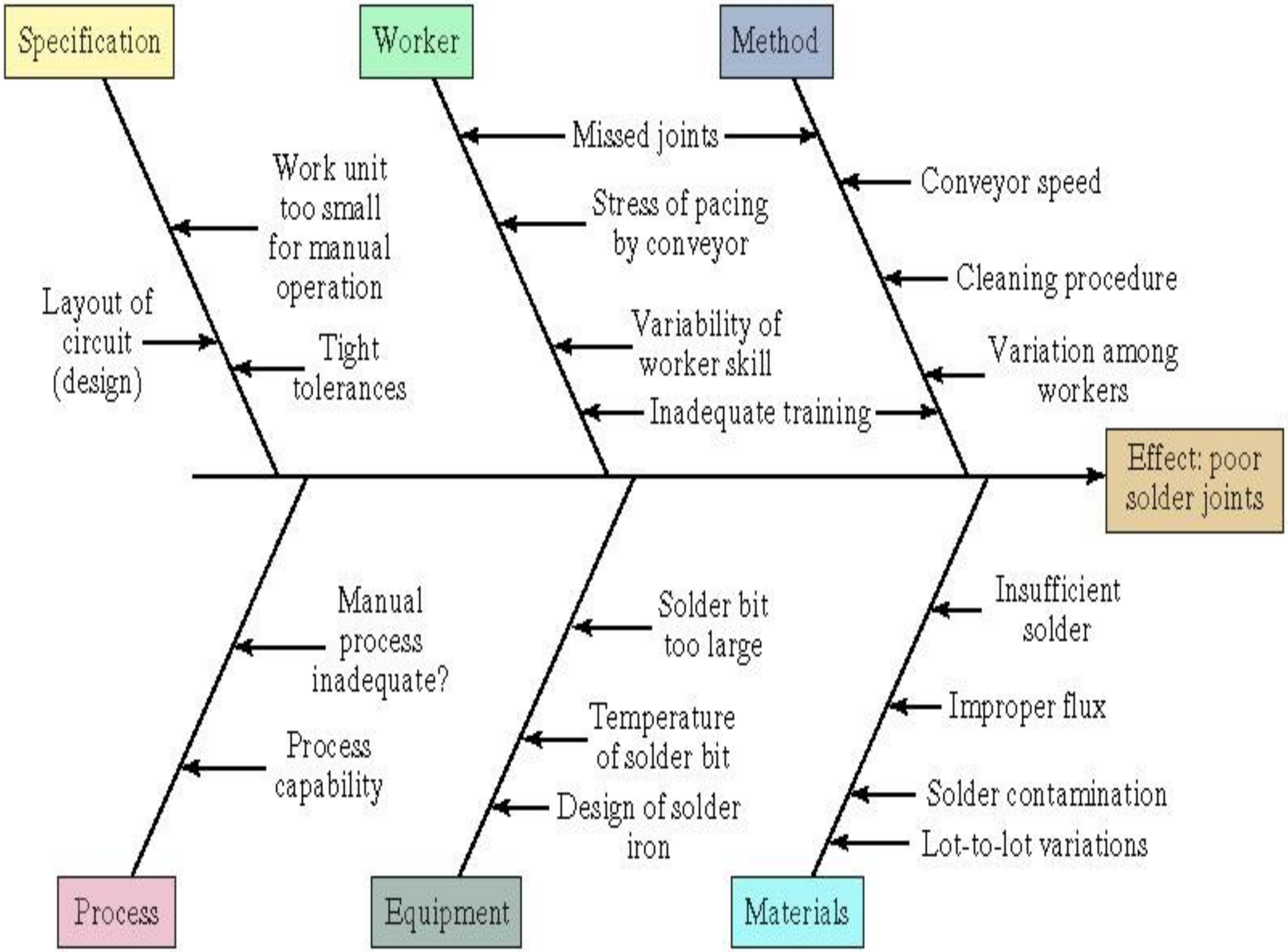
Figure : Pareto Diagram by Defective Items



4. Cause and effect diagrams

A graphical-tabular chart used to list and analyze the potential causes of a given problem

- Also known as a “fishbone diagram”
- Can be used to identify which causes are most consequential and how to take corrective action against them


Specification
Worker
Method
Missed joints
Work unit too small for manual operation
Stress of pacing by conveyor
Conveyor speed
Cleaning procedure
Layout of circuit (design)
Tight tolerances
Variability of worker skill
Variation among workers
Inadequate training
Effect: poor solder joints
Manual process inadequate?
Solder bit too large
Insufficient solder
Improper flux
Temperature of solder bit
Process capability
Design of solder iron
Solder contamination
Lot-to-lot variations
Process
Equipment
Materials

5. Scatter Diagram

Measuring Relationships between Variables. They are often used as follow-ups to a cause-and effect analysis to determine whether a stated cause truly does impact the quality characteristic.

The scatter diagram is a plot of one variable against another variable on a graph in an effort to see if there is a relationship between the two variables.

Phases in preparing Scatter Diagram:

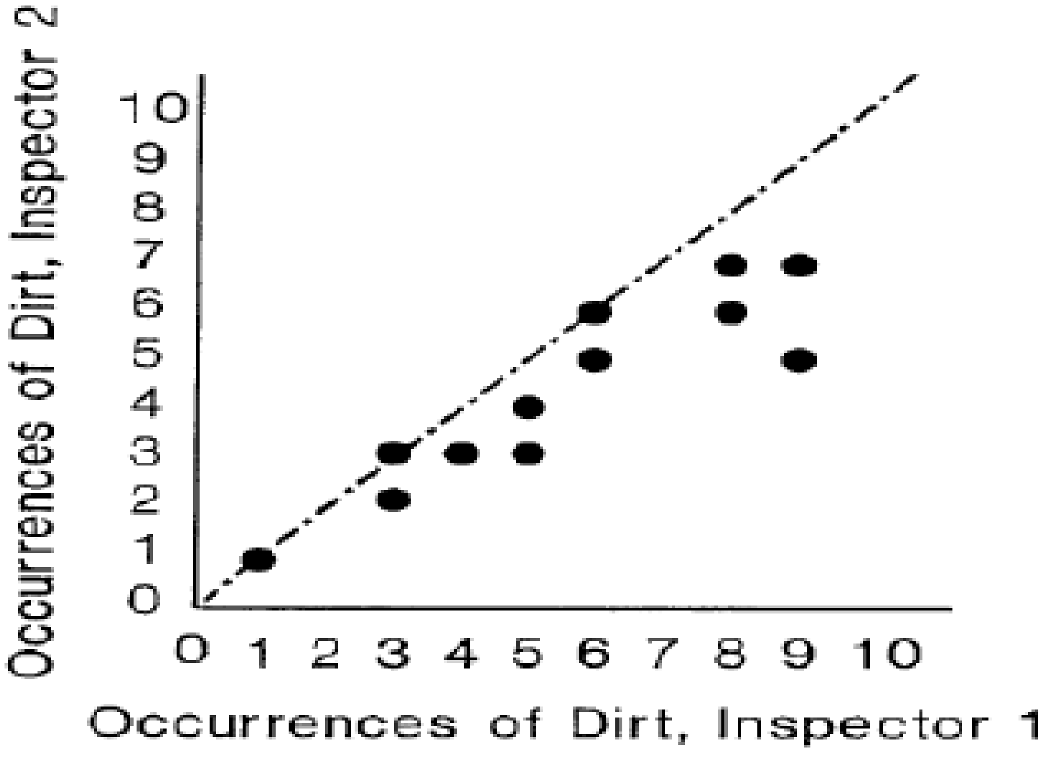
- **Analyze** – collect and draw the data points of two variables to be related in the x-y axis. Consider one variable as independent variable which should be in the x-axis and the other variable as dependent variable which should be in the y-axis. From the data points, try to investigate the relationship between the variables (whether it is positive or negative).
- **Improve (Implement**) – based on the relationship between the two variables, propose a solution for improvement.

It is used:

- To study and identify the possible relationship between the changes observed in two different sets of variables.
- It immediately provides a visual means to test the strength of a potential relationship.
- Provides a good follow-up to a Cause-and-Effect diagram to find out if there is more than just a consensus connection between causes and effects

- Does not predict cause-and-effect relationships. Only shows the strength of the relationship between two variables. The stronger the relationship, the greater the likelihood that change in one variable will affect change in another variable.
- An option is to stratify data sets to check for other sources of variation other than the effect of one variable to the other.

For example, in an attempt to standardize inspection, several pieces were inspected for evidence of dirt by two inspectors. The occurrences of dirt found by each of the two inspector on the same pieces is plotted in the figure. Note that the two inspectors are not in agreement, inspector 2 found fewer indication of dirt than does inspector 1. Training should be undertaken to bring the two inspectors into agreement to give the data credibility.



Scatter diagram on inspection data

6.Theory of control charts

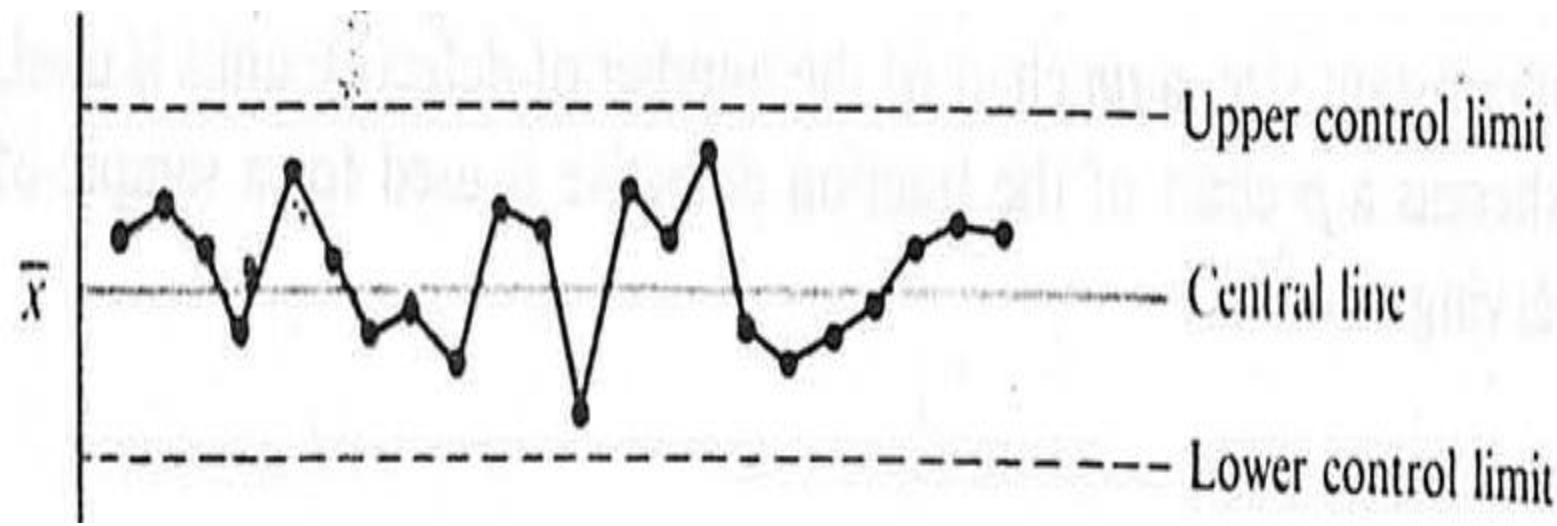
A control chart was first proposed in 1924 by W.A Shewhart, who belonged to the Bell telephone laboratories, with a view to eliminate an abnormal variation by distinguishing variations due to assignable causes from those due to chance causes.

A Control chart is a graphical method for displaying control results and evaluating a measurement procedure is in control or out-of-control.

A control chart consists of:

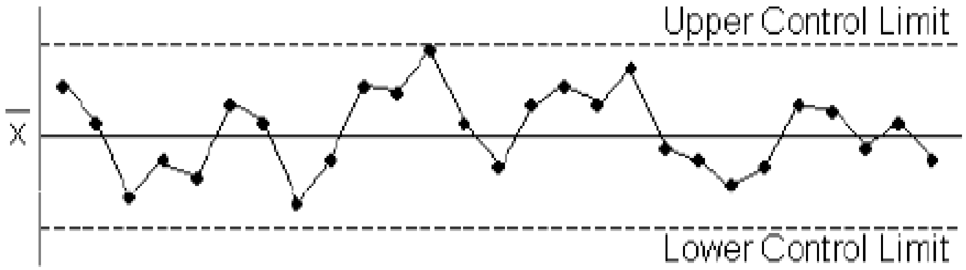
- A central line
- Upper control limit
- Lower control limit and

Characteristic values plotted on the chart which represent the state of a process.

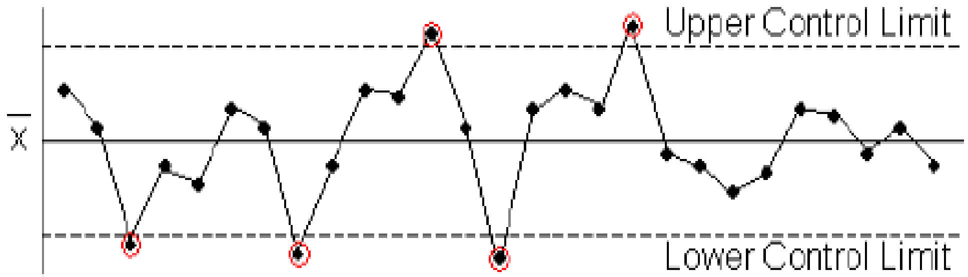


Upper control limit
$\bar{x}$
Central line
Lower control limit

If all these values are plotted within the control limits without any particular tendency, the process is regarded as being in the controlled state, however, otherwise it is out of control.



In - Control



Out of Control

Uses of Control charts

The main uses of control charts are:

1. It is a proven technique for improving productivity.

2. It is effective in defect prevention.

3. It prevents unnecessary process adjustments.

4. It provides diagnostic information.

5. It provides information about process capability.

7. Flow chart

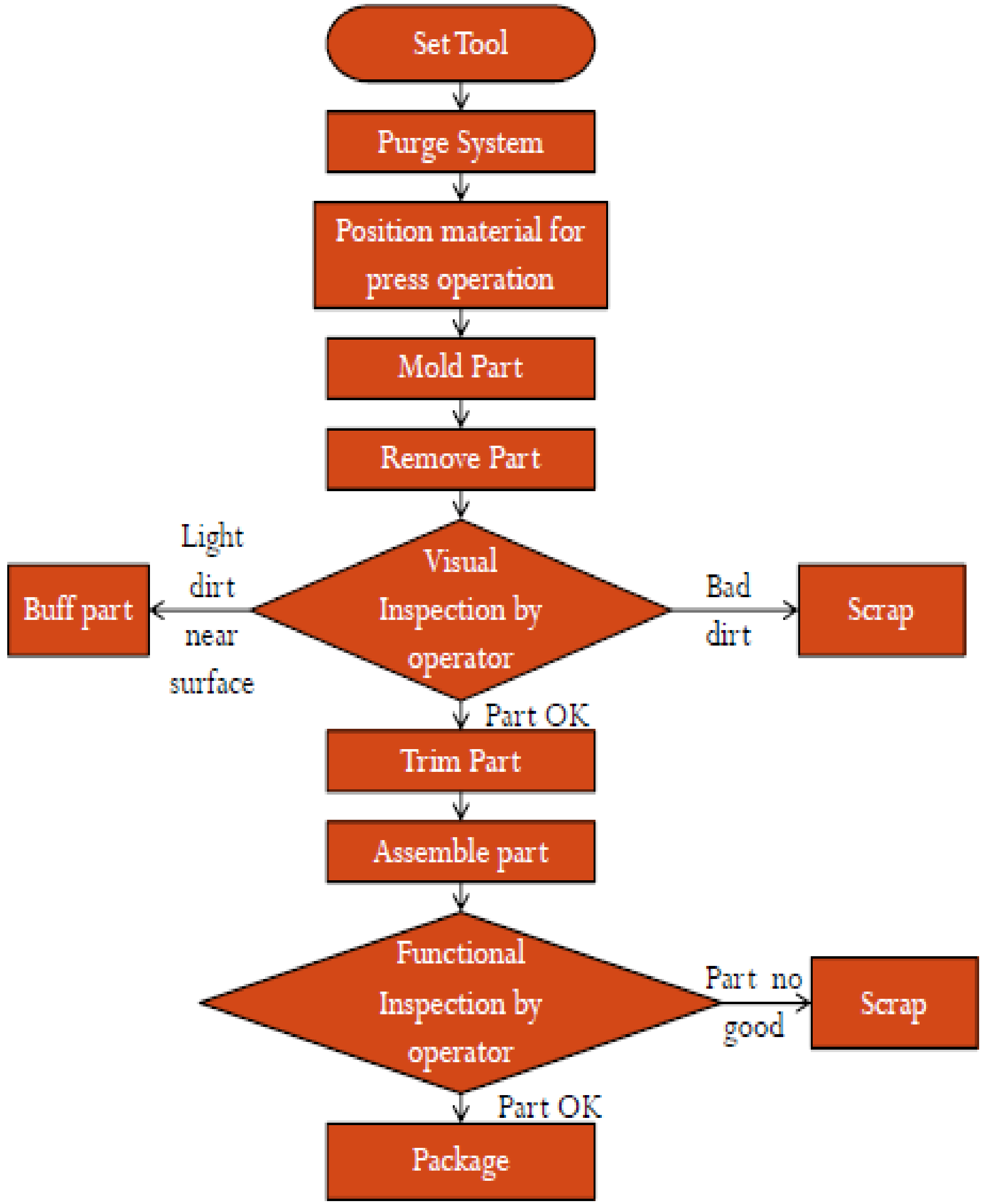
- It is picturing the process.
- It uses to allow a team to identify the actual flow or sequence of events in a process that any product or service follows.

Phases

- Define – this phase identifying which process to study and draw the process flow of the current process.
- Measure – identifying problems in the current flow chart
- Improve – Improving the flow chart to solve the identified problem

To Draw a Flow Chart you need to know:

- The Process
- Each activities in the process
- The precedence of the activities in the process
- The characteristics of the activities
- The decisions to be made, the data used and types of documents and activities in the process.



Flow chart of the molding process