β MODULE

D — DEFINE

- Launch Team
- Establish Charter
- Plan Project
- Gather the Voice of the Customer
- Plan for Change

M — MEASURE

- Document the Process
- Collect Baseline data
- Narrow project focus

A — ANALYZE

- Analyze Data
- Identify Root Cause
- Identify and Remove Wastes

I — IMPROVE

- Generate Solutions
- Evaluate Solutions
- Optimize Solutions
- Pilot
- Plan and implement

C — CONTROL

- Control the Process
- Validate project benefits

THE MAIN FOCUS OF β MODULE

- Introduction to Root Cause Analysis
- Data Driven Techniques
- Process focused techniques
- Validating root cause

Y= f (X)

To get results, should we focus our behavior on the Y or X ?

Y	Xi…..Xn
Dependent	Independent
Output	Input
Effect	Cause
Response	Factor

DEFINATION OF ROOT CAUSE

Root Cause

- Root Cause or a system of root causes is the underlying reason of a problem which has the following two characteristics – *PARADIES*
  - It is actionable
  - Once actioned, it shall prevent the occurrence of a problem.
- Root Cause is a MIN process. Something that is either missing or incomplete or not followed – *IVAN FANTIN*

M - Missing

I – Incomplete

N – Not followed.

What is 5 Whys Analysis?

The 5 Why Root Cause Analysis is a technique used to analyze any problem by repeatedly asking the question “Why”, which leads to the root cause of a problem. This lends a structured approach to help managers solve business problems. The 5 Why’s is an iterative process used to analyze the cause and effect relationships of a business problem.

Owing to its effective and lean nature, this **Root Cause Failure Analysis (RCFA)** Tool has gained popularity.

EXAMPLE OF WHY-WHY Analysis

Why ? 01 Because the account head didn't approve the deliverables on time

Why ? 02 Because the account head was absent on an unplanned leave

Why ? 03 Because a substitute account manager was not assigned for the approvals

Why ? 04 Because we had not predicted emergency leaves for the account head

Why ? 05 Because we had not predicted such situations hence not made rules around emergency planning

EXAMPLE OF WHY-WHY Analysis

An example of a problem is: The vehicle will not start.

1. *Why?* – The <u>battery</u> is dead. (First why)
2. *Why?* – The <u>alternator</u> is not functioning. (Second why)
3. *Why?* – The alternator <u>belt</u> has broken. (Third why)
4. *Why?* – The alternator belt was well beyond its useful service life and not replaced. (Fourth why)
5. *Why?* – The vehicle was not maintained according to the recommended service schedule. **(Fifth why, a root cause)**

List all Potential Causes / X's

Utilize one of the below tools to list all the potential causes

- Cause and Effect / Fishbone / Ishikawa Diagram
- Affinity Diagram

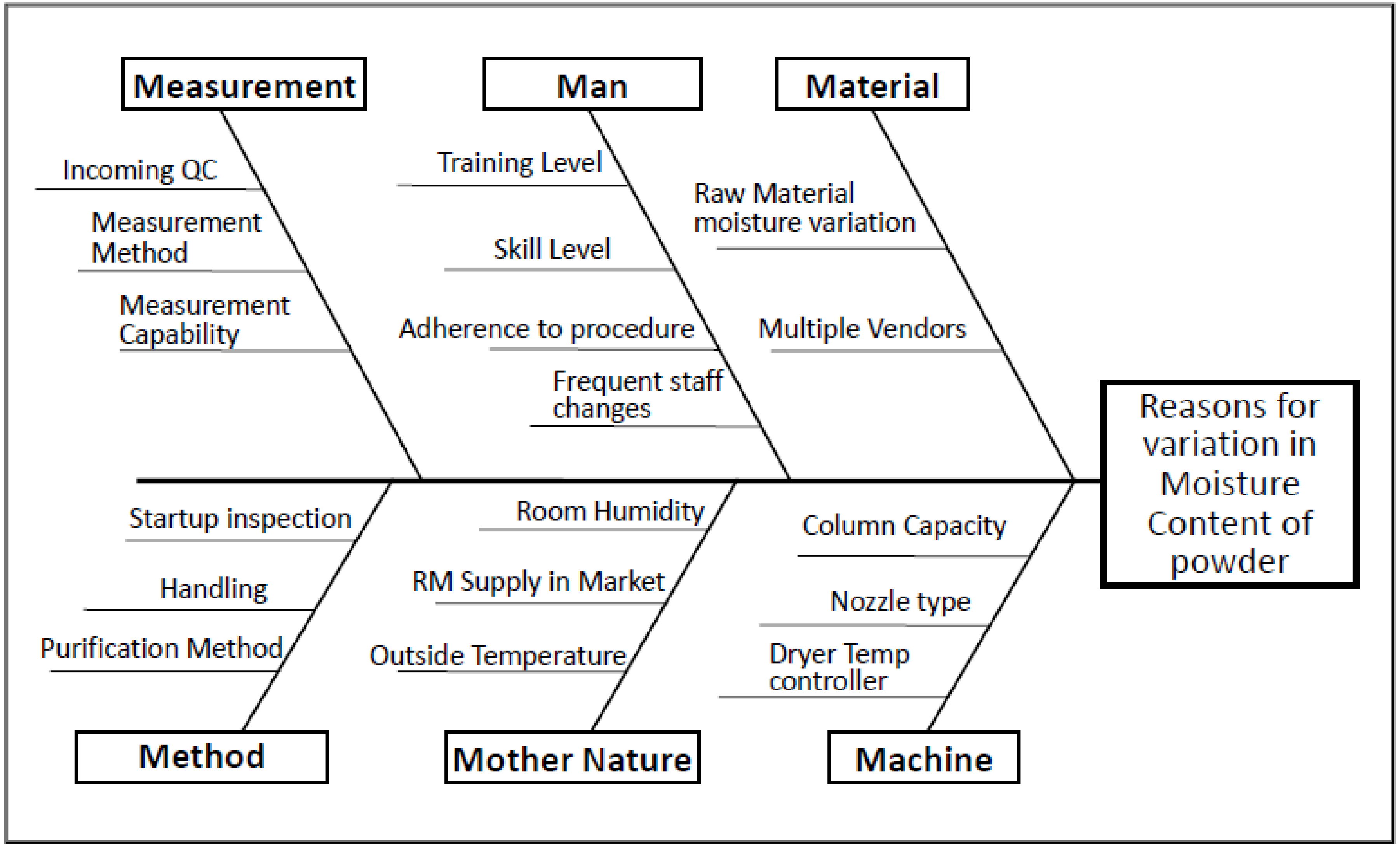
Cause & Effect / Fishbone / Ishikawa Diagram

- Helps identify potential causes to a problem in a pictorial format
- Format appears like the skeleton of a fish
- Developed by Kaoru Ishikawa in the 1960s, who pioneered quality management processes in the Kawasaki shipyards
- Causes in a typical diagram are normally grouped into categories, the main ones of which are:
  - The 6 Ms: Men/people, machines, methods, materials, measurement, mother nature (environment)
  - 4 Ps – Places, Procedures, People, Politics
  - 4 Ss – Surroundings, Suppliers, Systems, Skills
- Steps involved
  1. Identify the problem
  2. Work out the categories to be used
  3. Brainstorm the causes leading to the problem
  4. Analyze the diagram

List all Potential Causes / X's

EXAMPLE

List all Potential Causes / X's



Affinity Diagram

Is a tool used to generate ideas related to issues and then organizing these ideas into logical categories basis their underlying similarity. Affinity diagrams are useful for analysis of quality problems, defect data, customer complaints, survey results, etc. They can be used in conjunction with other techniques such as cause and effect diagrams

Steps involved in Affinity Diagram

1. Generate ideas: Brainstorming is immensely helpful in generating ideas
2. Display ideas: Write the ideas on small pieces of paper (Post-its work very well).
3. Sort ideas into groups: To put an idea into a category a person simply picks up the Post-it and moves it.
4. Create header cards: The final groupings are then reviewed and discussed by the team. Usually, the grouping of ideas helps the team to develop a coherent plan.
5. Draw finished diagram

Affinity Diagram

List all Potential Causes / X’s

Affinity Diagram

List all Potential Causes / X's

➡ **Problem: What are some of the ways to reduce cycle time for process A?**

Step 1: Generate or Create ideas

→ In the very first stage of this activity, we need to generate_ideas by using the Brainstorming process.

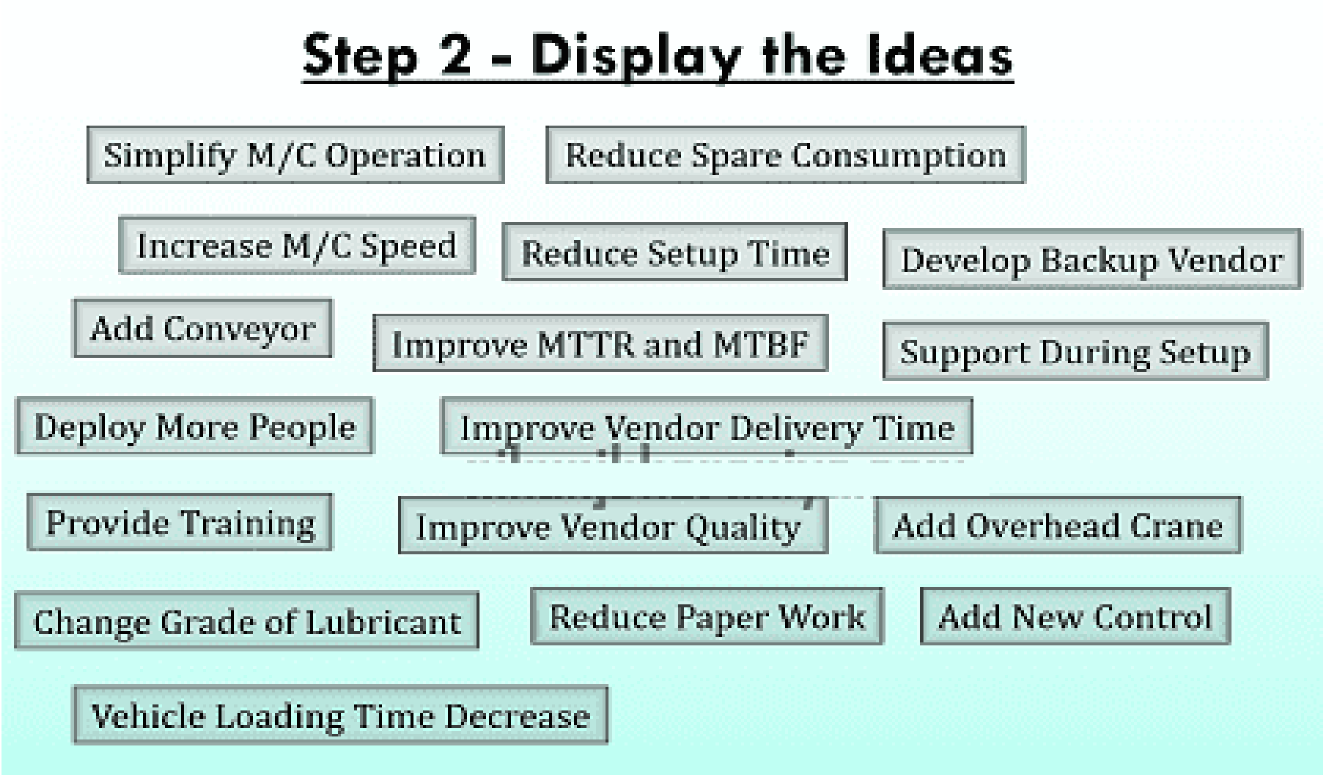
Step 2 - Display ideas

Affinity Diagram

List all Potential Causes / X's

→ After generating the_ideas note down in sticky note or in cards and display on a chart, wall, table, or board in random order.

→ For your better understanding refer to the below picture.



Affinity Diagram

List all Potential Causes / X's

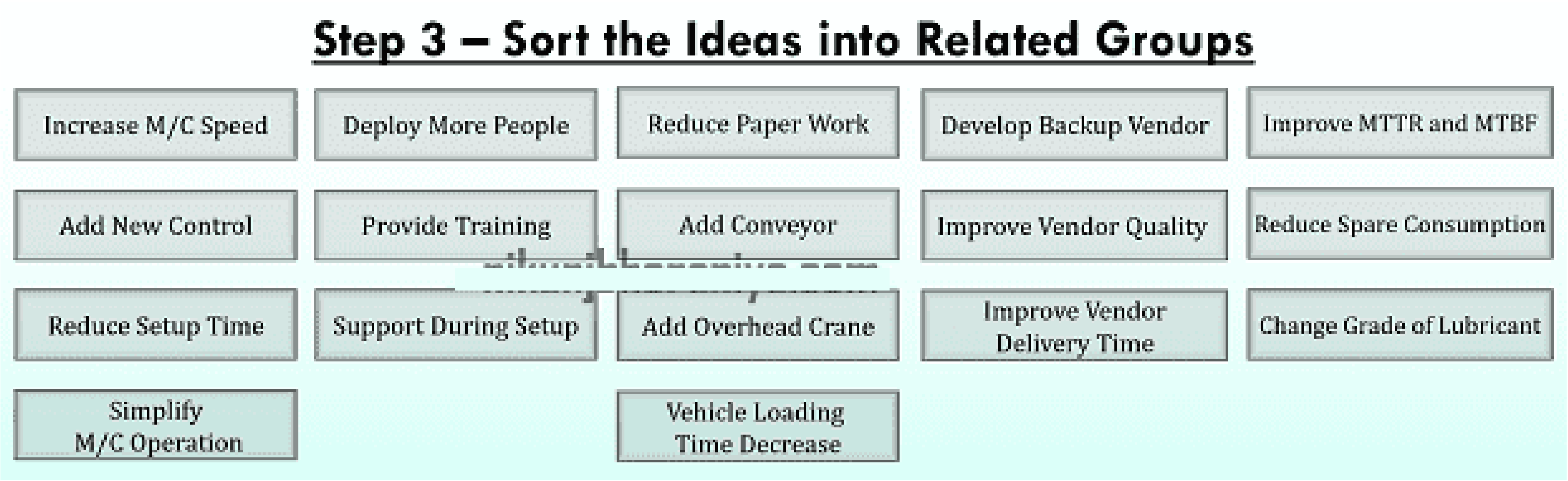
Step 3 - Sort the Ideas into Related Groups

→ Now in the 3rd step, we need to sort_ideas as per the relevant group.

→ We can start this activity by looking for two_ideas that seem related in some way.

→ Place them together in a column.

→ For better understanding refer to the below picture.



Step 3 - Sort the Ideas into Related Groups

Affinity Diagram

Group_01:
→ Increase M/C Speed
→ Add New Control
→ Reduce setup time
→ Simplify M/C Operation

Group_02:
→ Deploy More People
→ Provide Training
→ Support During Setup

Group_03:
→ Reduce paperwork
→ Add Conveyor
→ Add Overhead Crane
→ Vehicle Loading Time Decrease

Group_04:
→ Improve Communication with Vendor
→ Develop Backup Vendor
→ Improve Vendor Quality
→ Improve Vendor Delivery Time

Group_05:
→ Change Grade of Lubricant
→ Improve MTTR and MTBF
→ Reduce Spare Consumption

List all Potential Causes / X's

Affinity Diagram

List all Potential Causes / X's

Step 4 – Create Header Cards for the Groups

→ A header is an idea that includes and captures the essential relationships between the_ideas contained in a group of cards.

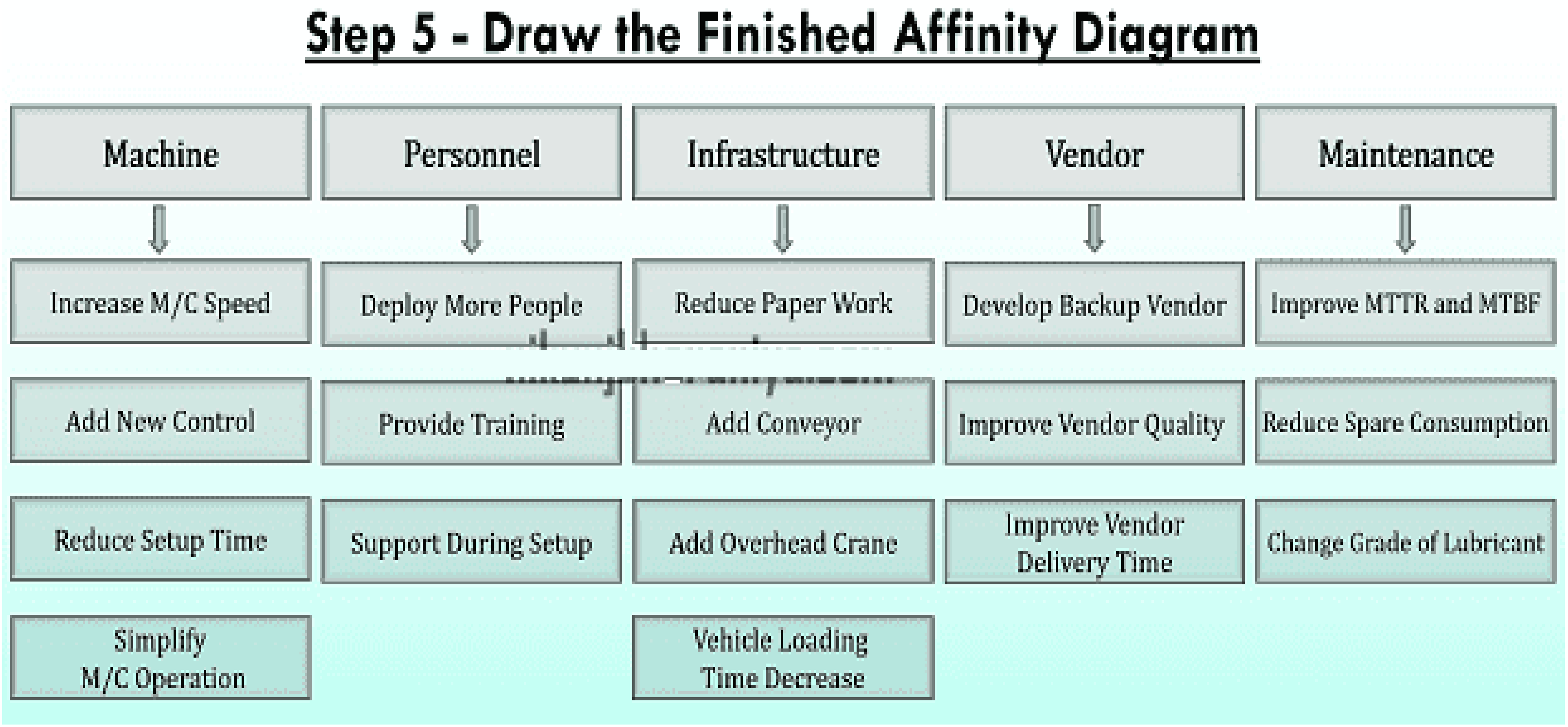
→ We can see the header for the relevant activity in the below picture.

→ Now Based on the group discussion we can create the below headers for the above-mentioned groups.

➡ Group_01 = **Machine**
➡ Group_02 = **Personnel**
➡ Group_03 = **Infrastructure**
➡ Group_04 = **Vendor**
➡ Group_05 = **Maintenance**

Affinity Diagram

List all Potential Causes / X's



Affinity Diagram

List all Potential Causes / X's

Benefits of the Affinity Diagram Tool and Process

→ The affinity diagram process is simple and cost-effective.

→ The thinking ability of members will increase.

→ Innovative ideas will come during brainstorming.

→ This is a transparent and systematic tool.

→ This is a very simple and effective tool that everyone can use easily.

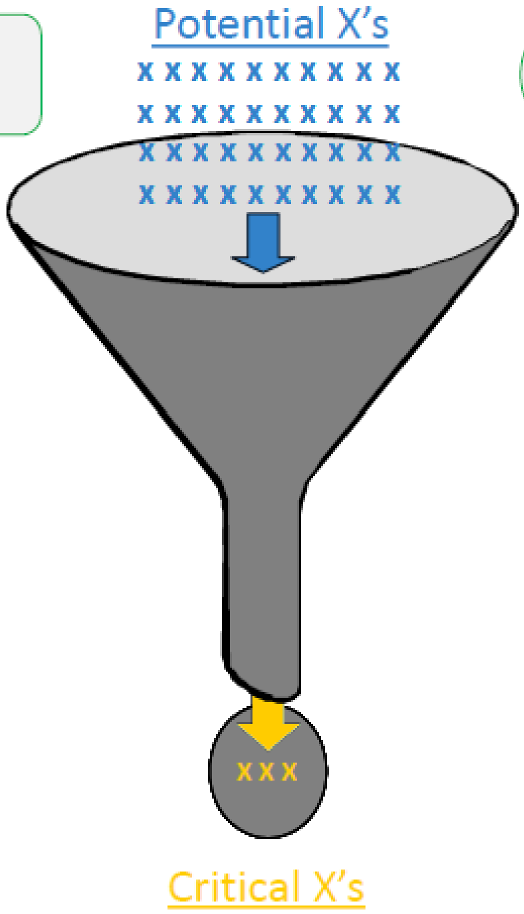
Identify Critical Causes / X's

A **Graphical Tools**

1. Pareto Plot
2. Scatter Plot

B **Process Map Based Tools**

1. FMEA
2. VA/NVA Analysis



C **Statistical Tools**

1. 2 Proportion Test
2. Chi square Test
3. 2 Sample T Test
4. ANOVA
5. Correlation and Regression

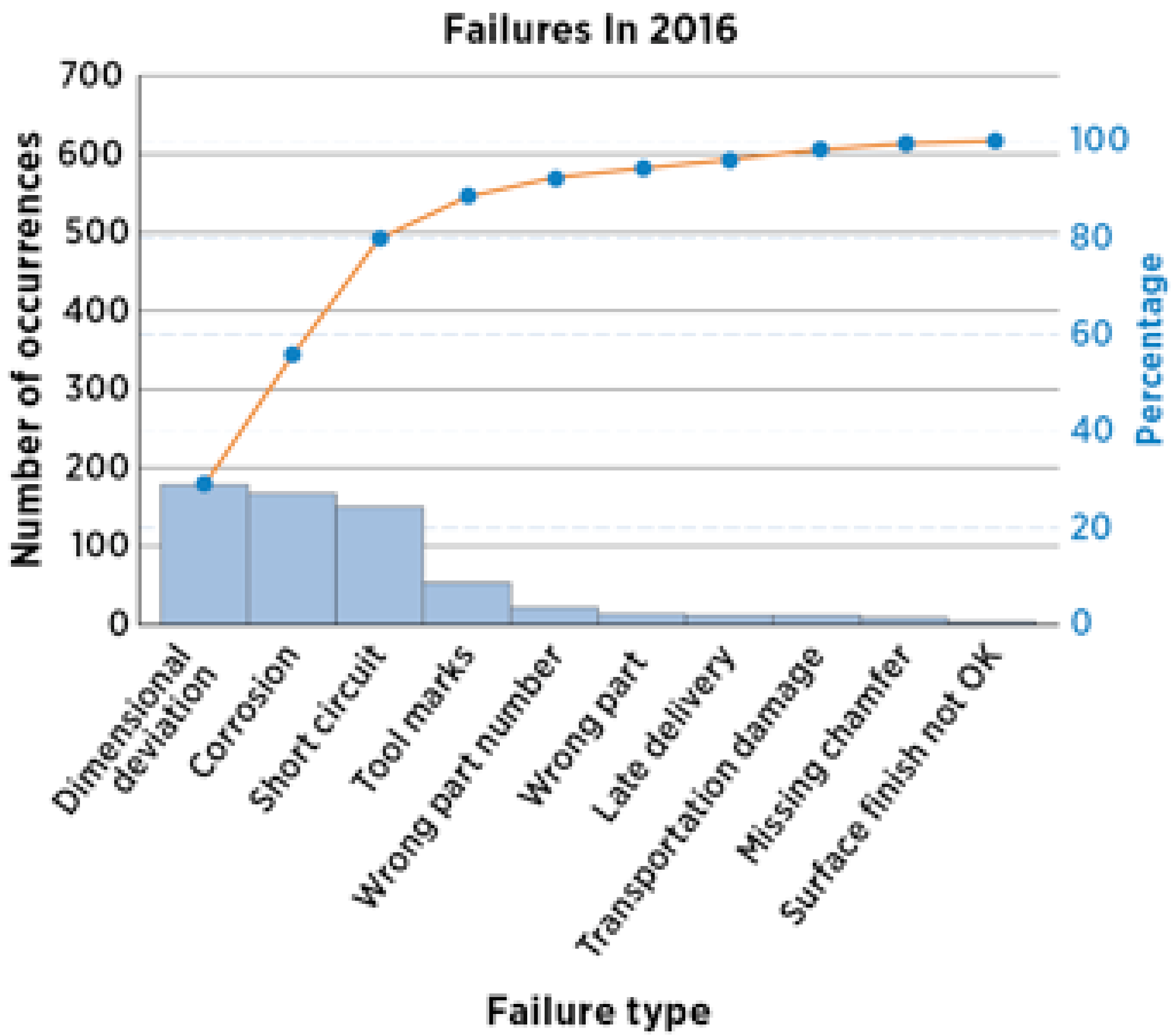
Pareto Analysis

- ✓ Pareto analysis is a bar graph used to organize data in such a way that shows what major X's influence Y. It is frequently referred to as **"the search for significance"**
- ✓ The basis for building a Pareto is the **80/20 rule.** Typically, approximately 80% of the problem(s) result from approximately 20% of the causes
- ✓ An outcome of Pareto's Analysis is Pareto chart

A Graphical Tools

Identify Critical Causes / X's

Pareto Analysis



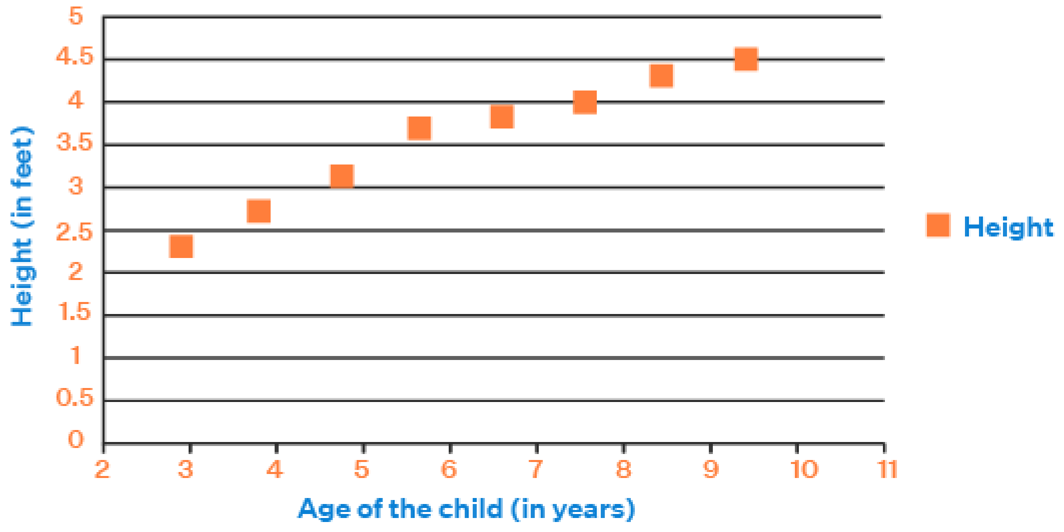
Pareto Examples

- ✓ 80% of your phone calls go to 20% of the names on your list
- ✓ 80% of the meals in a restaurant come from 20% of the menu
- ✓ 80% of sales come from top 20% of the customers
- ✓ 80% of sales come from 20% of products
- ✓ 80% of complaints are received for 20% of service request types

Scatter Plot

What Is a Scatter Plot?

A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

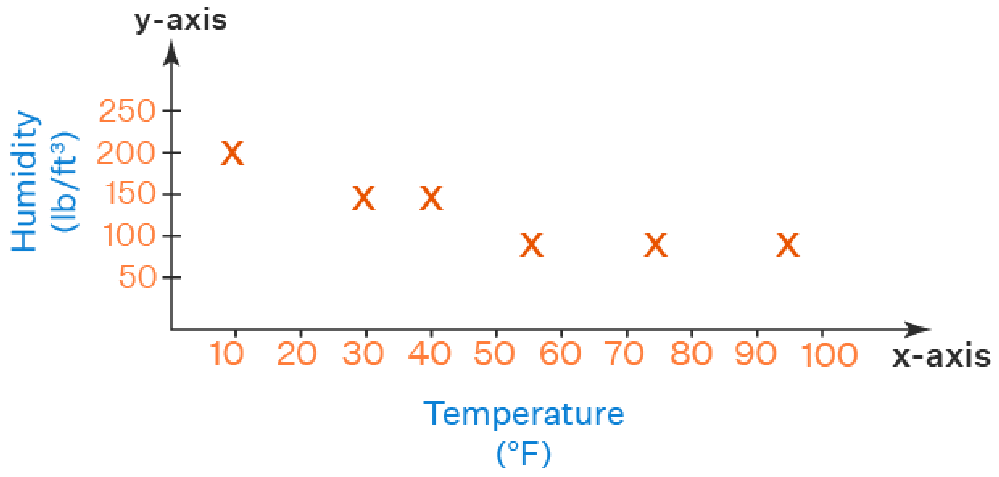


Age of the Child	Height
3	2.3
4	2.7
5	3.1
6	3.6
7	3.8
8	4
9	4.3
10	4.5

How to Construct a Scatter Plot?

There are three simple steps to plot a scatter plot

- **STEP I:** Identify the x-axis and y-axis for the scatter plot.
- **STEP II:** Define the scale for each of the axes.
- **STEP III:** Plot the points based on their values.



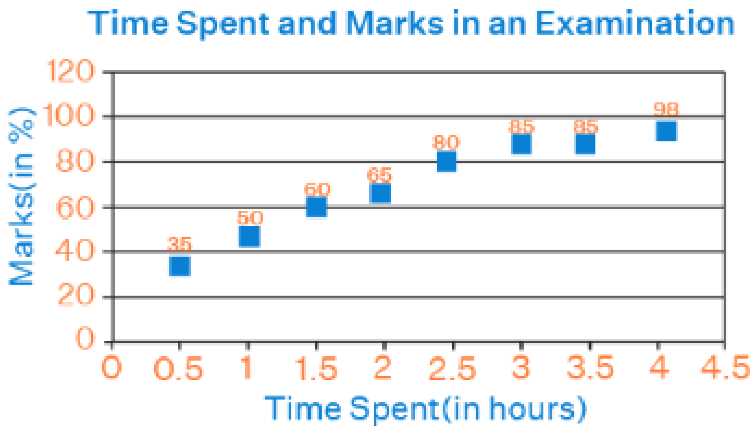
Types of Scatter Plot

A scatter plot helps find the relationship between two variables. This relationship is referred to as a correlation. Based on the correlation, scatter plots can be classified as follows.

- Scatter Plot for Positive Correlation
- Scatter Plot for Negative Correlation
- Scatter Plot for Null Correlation

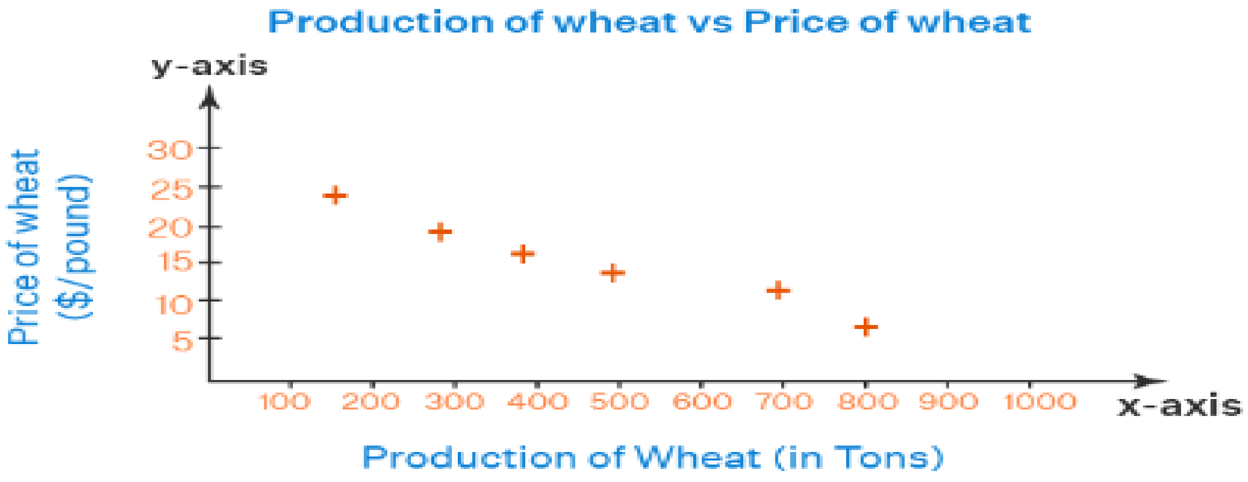
Scatter Plot for Positive Correlation

A scatter plot with increasing values of both variables can be said to have a positive correlation. The scatter plot for the relationship between the time spent studying for an examination and the marks scored can be referred to as having a positive correlation.



Scatter Plot for Negative Correlation

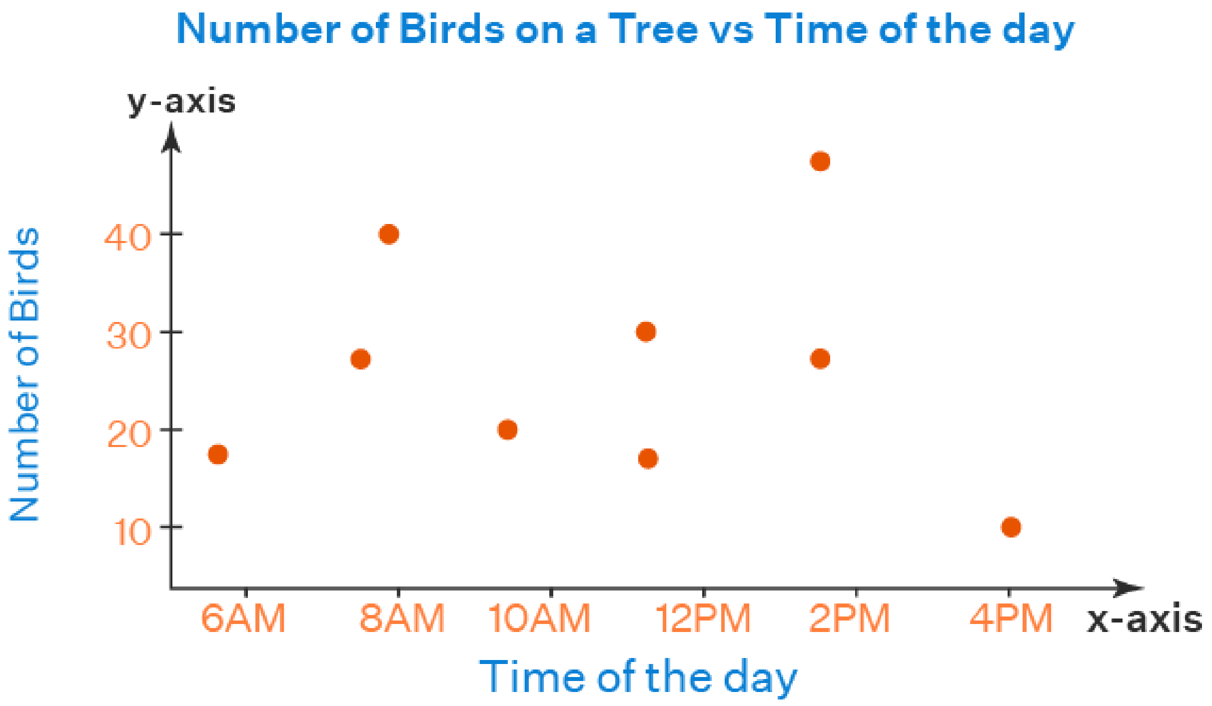
A scatter plot with an increasing value of one variable and a decreasing value for another variable can be said to have a negative correlation. Observe the below image of negative scatter plot depicting the amount of production of wheat against the respective price of wheat.



Scatter Plot

Scatter Plot for Null Correlation

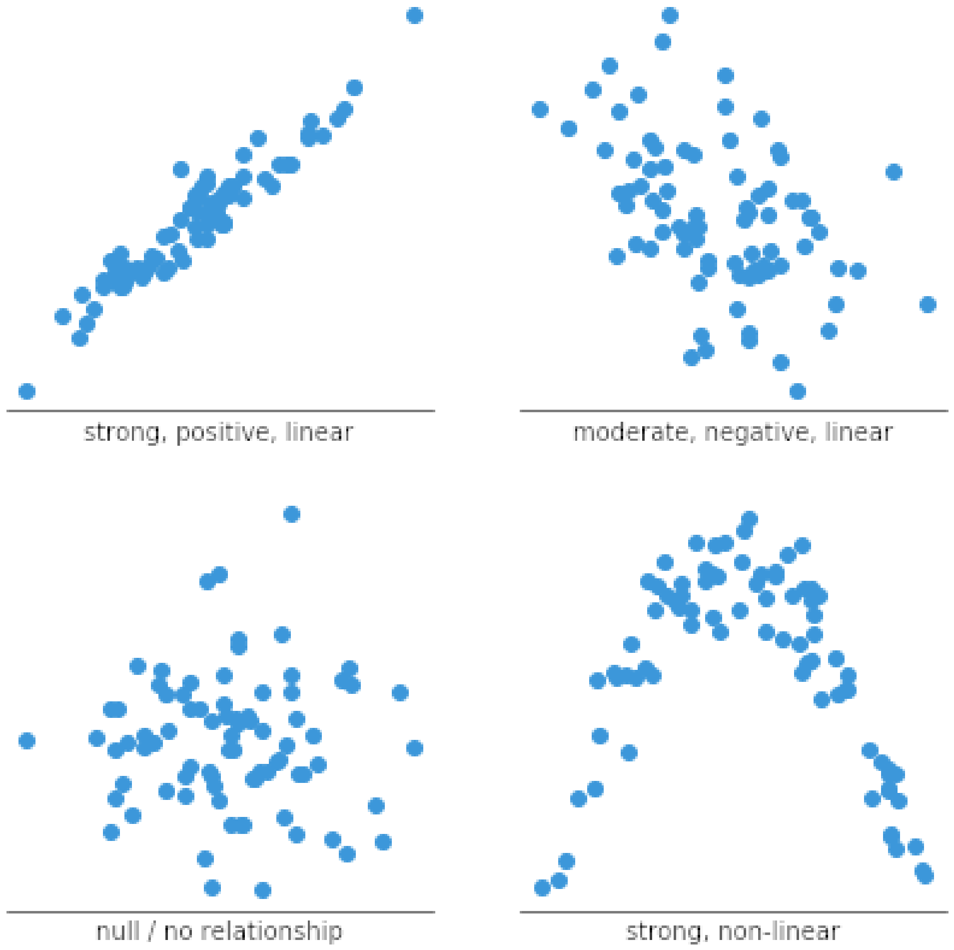
A scatter plot with no clear increasing or decreasing trend in the values of the variables is said to have no correlation. Here the points are distributed randomly across the graph. For example, the data for the number of birds on a tree at different times of the day does not show any correlation. Observe the below scatter plot showing the number of birds on a tree versus time of the day.



A **Graphical Tools**

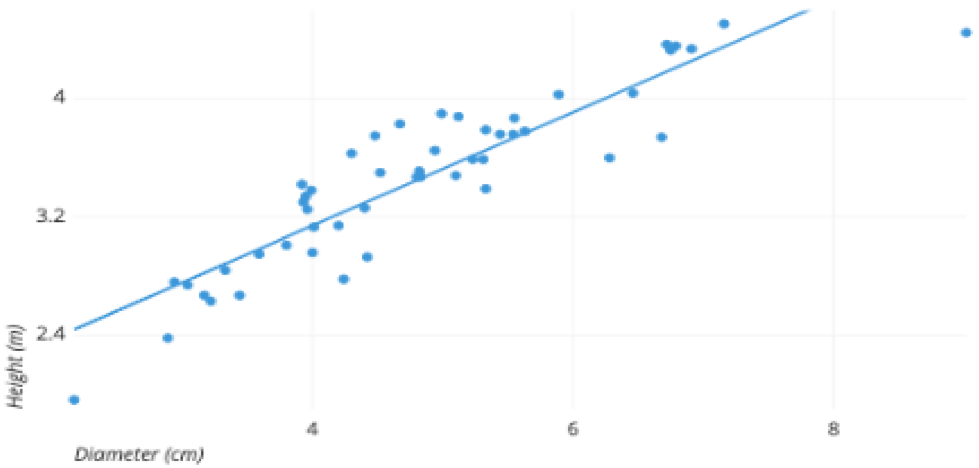
Identify Critical Causes / X's

Scatter Plot



Add a trend line

When a scatter plot is used to look at a predictive or correlational relationship between variables, it is common to add a trend line to the plot showing the mathematically best fit to the data. This can provide an additional signal as to how strong the relationship between the two variables is, and if there are any unusual points that are affecting the computation of the trend line.



B Process Map Based Tools

Identify Critical Causes / X's

1. VA/NVA Analysis
2. FMEA

Value Added Activities must satisfy the following three criteria:

1. Work that the customer is willing to pay for
2. Work that physically transforms the product (or document/information)
3. Work that is done right the first time

Use the acronym CPR to remember: **C**ustomer pays for it, **P**hysically transforms the product, **R**ight the first time.

Non value-Added Step:

- Is not essential to produce output.
- Does not add value to the output.

Non-Value-Added Steps can be further classified

1. Required NVA

Business necessity (e.g. accounting)

Employee necessity (e.g. payroll)

Process necessity (e.g. inspection)

2. Pure Waste/ NVA

VA and NVA (Lean)

VA/NVA analysis procedure

Step 1: Map the process like a flowchart detailing each activity, arrow and decision.

Step 2: For each arrow, box, and diamond, list its function and the time spent (in minutes, hours, days) on the value-added check list

Step 3: Now become the customer. Step into their shoes and ask the following questions:

- Is the customer not willing to pay for it?
- Is this inspection, testing, or checking ?
- Is this just "fix it" error correction work or waste?

Step 4: If the answer to any of these questions is "yes", then the step may be non-value-added.

Step 5: How can activities and delays be eliminated, simplified, combined, or reorganized to provide a faster, higher quality flow through the process?

VA and NVA (Lean)

WASTE

Mnemonics to Remember

WORMPITS	
W	Waiting
O	Over Production
R	Rejects
M	Motion
P	Over Processing
I	Inventory
T	Transportation
S	Skillset

TIMWOODS	
T	Transportation
I	Inventory
M	Motion
W	Waiting
O	Over Production
O	Over Processing
D	Defects
S	Skillset

DOWNTIME	
D	Defects
O	Over Production
W	Waiting
N	Non-utilized Talent
T	Transportation
I	Inventory
M	Motion
E	Extra-Processing

VA and NVA (Lean)

Waiting

Waiting is non-productive time due to lack of material, people, or equipment. Waste of Waiting is the cost of an idle resource.

Examples are:

- ✓Processing once each month instead of as the work comes in
- ✓Showing up on time for a meeting that starts late
- ✓Delayed work due to lack of communication from another internal group

W A S T E

B Process Map Based Tools

VA and NVA (Lean)

Identify Critical Causes / X's

Overproduction

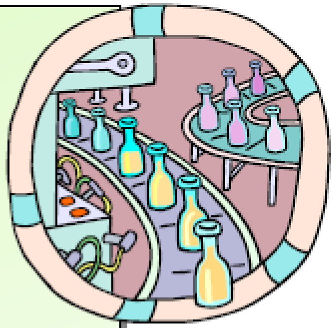
- Overproduction is producing more than the next step needs or more than the customer buys.
- It may be the worst form of waste because it contributes to all the others.
- Waste of Overproduction relates to the excessive accumulation of work-in-process (WIP) or finished goods inventory.

Examples are:

✓Preparing extra reports

✓Adding more features than required

✓Producing parts/components more than sales demand



W
A
S
T
E

VA and NVA (Lean)

Over-processing

- Over-processing is tasks, activities and materials that don't add value.
- Can be caused by poor product or tool design as well as from not understanding what the customer wants.

Examples are:

✓Reports that contain more information than the customer wants or needs

✓Communications, reports, emails, contracts, etc. that contain more than the necessary points (briefer is better)

✓Voice mails that are too long

W
A
S
T
E

VA and NVA (Lean)

Defects/ Rework

Waste of Correction includes the waste of handling and fixing mistakes. This is common in both manufacturing and transactional settings.

Examples are:

✓Incorrect data entry

✓Paying the wrong vendor

✓Misspelled words in communications

✓Non-conforming output

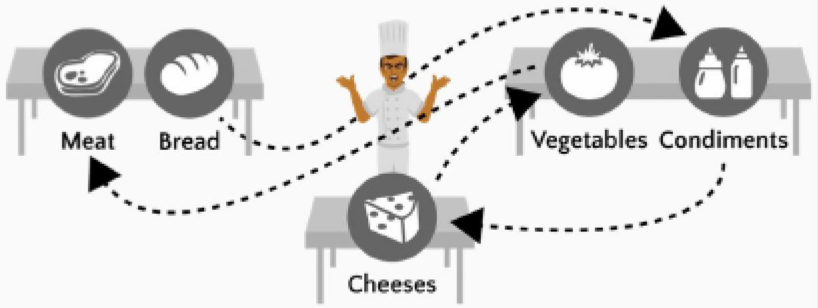
✓Materials discarded during production

W
A
S
T
E

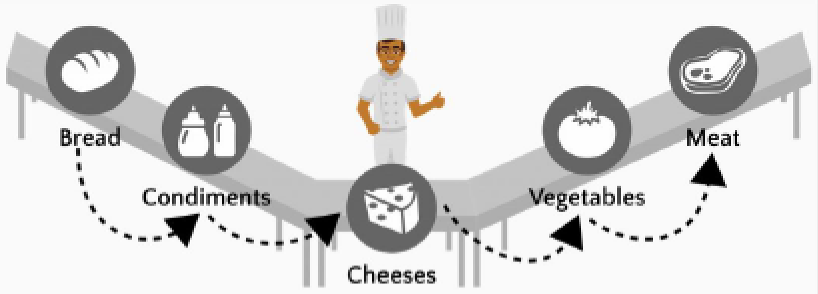
VA and NVA (Lean)

Motion

- Motion is the unnecessary movement of people and equipment.
- This includes looking for things like documents or parts as well as movement that is straining
- Waste of Motion examines how people move to ensure that value is added.



Before Waste is removed in a kitchen (Sandwich making process example).



After Waste is removed in a kitchen (Sandwich making process example).

Examples are:

✓Task-Switching

✓Movement of people in a warehouse/factory floor

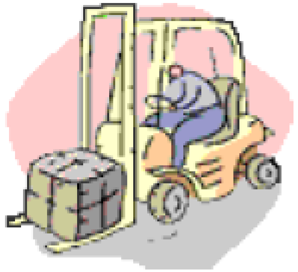
✓Steps required to get a task done/Navigation

W
A
S
T
E

VA and NVA (Lean)

Transportation

- Transportation is the unnecessary movement of material, goods and information
- Steps in a process should be located close to each other so movement is minimized.



Examples are:

✓Hand-Off's

✓ Physical travel for work (vs Video conference)

✓Moving files from desk to desk

W
A
S
T
E

B Process Map Based Tools

VA and NVA (Lean)

Identify Critical Causes / X's

Inventory

Inventory is the liability of materials that are bought, invested in and not immediately sold or used. Waste of Inventory is identical to overproduction except that it refers to the waste of acquiring raw material before the exact moment that it is needed.

Examples are:

✓Transactions not processed

✓Bench

✓Bigger "in box" than "out box"

✓Over-ordering raw materials

W
A
S
T
E

FMEA

Risk Analysis

What is Risk?

Risk can be defined as the likelihood of occurrence of an undesirable event combined with the magnitude of its impact

What is Risk Assessment?

Risk assessment is the determination of quantitative and qualitative value of risk related to a concrete situation and a recognized threat (also called hazard).

What does Risk Analysis include?

Risk analysis includes

- ✓ Identification of risks and the magnitude of their potential impact
- ✓ Estimation of their likelihood of occurrence
- ✓ Estimation of their causes
- ✓ Risk evaluation and Development of risk mitigation plan

FMEA

Failure Modes and Effects Analysis (FMEA)

It is a structured approach to:

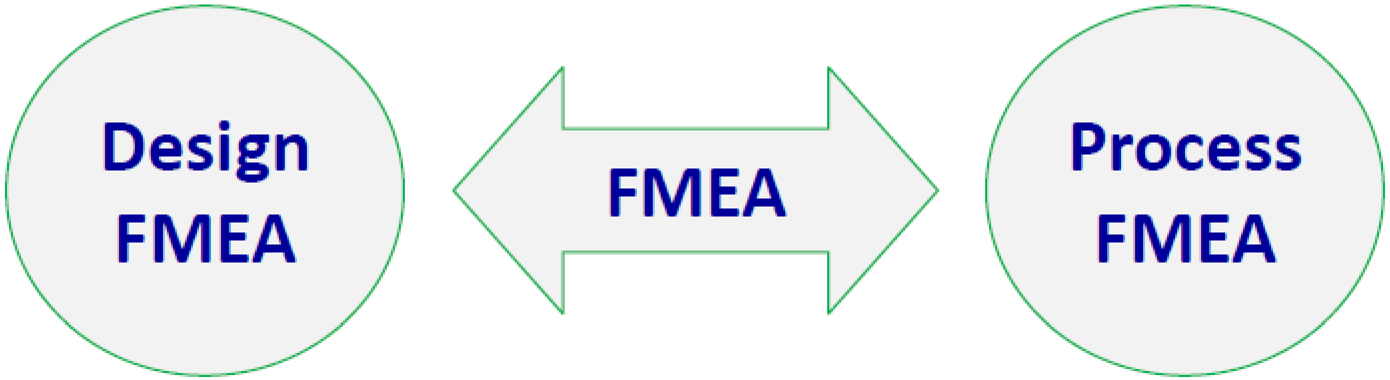
- Predict failures and prevent their occurrence in manufacturing and other functional areas which generate defects.
- Identify the ways in which a process can fail to meet critical customer requirements (Y).
- Estimate the **Severity, Occurrence and Detection** (SOD) of defects
- Evaluate the current Control Plan for preventing these failures from occurring and escaping to the customer.
- Prioritize the actions that should be taken to improve and control the process using a **Risk Priority Number** (RPN)

This enables us to evaluate current risks in the process and thereafter developing an action plan to mitigate risks

It is developed and maintained by multi-disciplinary (or cross functional) team typically led by the Green Belt/ Black Belt/Process Owner

FMEA

FMEA Types



During design, it is advantageous to know:

a) How and where customers will use product?

b) How customer may "abuse" product?

Performed to identify and address areas of potential risk within existing process.

Helps proactively manage risks in a process

FMEA

Definitions

Process step or product function: It indicates the item being analyzed. If the FMEA were based on a flow chart, it would be a step in the chart

Potential failure mode: It describes the way in which the process step or product function could possibly fail to satisfy its intended purpose. There may be more than one potential failures to a process step and each one should be listed separately

Potential effect of the failure: It describes the impact of the failure on the functionality of the product or process or what the customer or process owner will experience. The effect must be a very clear and explicit description of the impact of the noncompliance

Severity: The severity measures how critical or serious a potential failure can be on the product or process. If the failure is so serious that it can stop production, it is graded 10 and if it is very easy to correct, it is graded 1

FMEA

Definition

Potential causes of the failure: It describes the flaws of the design that may lead to a potential failure. Every conceivable reason for failure should be listed so that the corrective actions can be more promptly and cost effectively taken

Occurrence: The occurrence measures how often the failure is likely to happen. Here again, the likelihood of an occurrence is expressed in ranking from 1 to 10 (1 for rare and 10 for often)

Current control: Current controls are known preventive or detective controls that are currently being used in similar processes or products

Detection: How easy is the failure to detect? Detection is a measure of the ability to detect pre-emptively the failures. If the potential failure is easy to detect, the grade should be low (1 for very easy to detect and 10 for very hard)

FMEA

Severity Ranking Scale

A numerical measure of how serious is the effect of the failure.

	Effect	Design or Process: Customer Effect	Process: Manufacturing/Assembly Effect
10	Hazardous without warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation without warning.	Or may endanger operator (machine or assembly) without warning
9	Hazardous with warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation with warning.	Or may endanger operator (machine or assembly) with warning.
8	Very High:	Vehicle/item inoperable (loss of primary function).	Or 100% of product may have to be scrapped, or vehicle/item repaired in repair department with a repair time greater than one hour.
7	High	Vehicle/item operable, but at a reduced level of performance. Customer very dissatisfied.	Or product may have to be sorted and a portion (less than 100%) scrapped, or vehicle/item repaired in repair department with a repair time between a half-hour and an hour.
6	Moderate	Vehicle/item operable, but Comfort/Convenience item(s) inoperable. Customer dissatisfied.	Or a portion (less than 100%) of the product may have to be scrapped with no sorting, or vehicle/item repaired in repair department with a repair time less than a half-hour.
5	Low	Vehicle/item operable, but Comfort/Convenience item(s) operable at a reduced level of performance.	Or 100% of product may have to be reworked, or vehicle/item repaired off-line but does not go to repair department.
4	Very Low	Fit & Finish/Squeak & Rattle item does not conform. Defect noticed by most customers (greater than 75%).	Or product may have to be sorted, with no scrap, and a portion (less than 100%) reworked.
3	Minor	Fit & Finish/Squeak & Rattle item does not conform. Defect noticed by 50% of customers.	Or a portion (less than100%) of the product may have to be reworked, with no scrap, on-line but out-of-station.
2	Very minor	Fit & Finish/Squeak & Rattle item does not conform. Defect noticed by discriminating customers (less than 25%).	Or a portion (less than100%) of the product may have to be reworked, with no scrap, on-line and in-station.
1	None	No discernable effect	Or slight inconvenience to operation or operator, or no effect.

B Process Map Based Tools

FMEA

Identify Critical Causes / X's

Occurrence Ranking Scale

A numerical measure of how frequently the failure will happen

	Likelihood	Either ...	Or ...	Cpk
10	Very High:	More than 100 per thousand machines/items/ pieces	1 in 10 or less	<0.55
9	Persistent Failures	50 per thousand machines/items/pieces	1 in 20-50	0.55 to 0.78
8	High:	20 per thousand machines/items/pieces	1 in 50-100	0.78 to 0.86
7	Frequent Failures	10 per thousand machines/items/pieces	1 in 100-200	0.86 to 0.94
6	Moderate:	5 per thousand machines/items/pieces	1 in 200 -500	0.94 to 1.00
5	Occasional	2 per thousand machines/items/pieces	1 in 500-1000	1.00 to 1.10
4	Failures	1 per thousand machines/items/pieces	1 in 1000-2000	1.10 to 1.20
3	Low: Relatively	0.5 per thousand machines/items/pieces	1 in 2,000-10,000	1.20 to 1.30
2	Few Failures	0.1 per thousand machines/items/pieces	1 in 10,000-100,000	1.30 to 1.67
1	Remote: Failure is unlikely	Less than 0.010 per thousand machines/item/ pieces	1 in 100,000 or more	>=1.67

FMEA

Detection Ranking Scale

A numerical measure of how easy is it to detect a failure or cause of failure

	Detection	Criteria: Likelihood of Detection by PROCESS control	A	B	C	Suggested Range of Detection Methods
10	Absolute certainty of non-detection	Design Control will not and/or cannot detect a potential cause/mechanism and subsequent failure mode; or there is no Design Control.			X	Cannot detect or is not checked.
9	Controls will probably not detect	Very remote chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.			X	Control is achieved with indirect or random checks only.
8	Controls have poor chance of detection	**Remote chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.**			X	Control is achieved with visual inspection only.
7	Controls have poor chance of detection	**Very low chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.**			X	Control is achieved with double visual inspection only.
6	Controls may detect	Low chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.		X	X	Control is achieved with charting methods, such as SPC (Statistical Process Control)
5	Controls may detect	Moderate chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.		X		Control is based on variable gauging after parts have left the station, or Go/No Go gauging performed on 100% of the parts after parts have left the station.
4	Controls have a good chance to detect	Moderately high chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.	X	X		Error detection in subsequent operations OR gauging performed on setup and first-piece check (for set-up causes only).
3	Controls have a good chance to detect	High chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.	X	X		Error detection in-station, or error detection in subsequent operations by multiple layers of acceptance: supply, select, install, verify. Can't accept discrepant part.
2	Controls almost certain to detect	Very high chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.	X	X		Error detection in-station (automatic gauging with automatic stop feature.) Cannot accept discrepant part.
1	Controls certain to detect	Design Control will certainly detect a potential cause/mechanism and subsequent failure mode.	X			Discrepant parts cannot be made because item has been error-proof by process/product design.

*** Inspection Types: A= Error Proofed, B= Gauging, C= Manual Inspection**

FMEA

Risk Priority Numbers (RPN)

It is a numerical and relative "measure of overall risk" corresponding to a particular failure mechanism and is computed as follows

Risk Priority Number = Severity * Occurrence * Detection

RPN = S * O * D

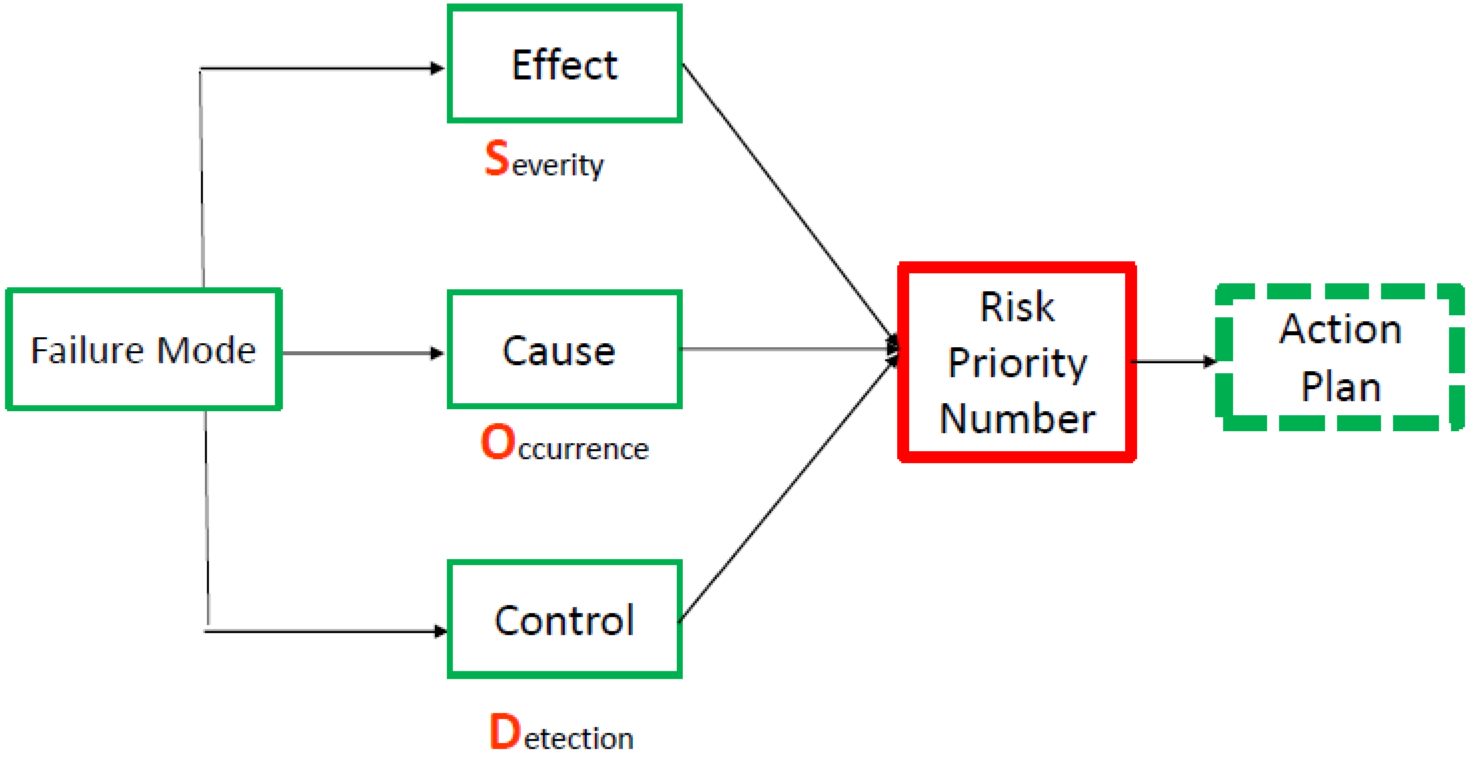
- The risk priority number (RPN) helps to rank the failures and establish their precedence for problem resolution considerations. It measures the process or product risk.
- The higher the RPN, higher the risk and hence more attention that a particular step of the process or that characteristic of the product should get
- RPN acts as a guideline and not a threshold for determining action priorities

B Process Map Based Tools

FMEA

Identify Critical Causes / X's

FMEA Concept



Action Plan

Preventive tasks are assigned to the task owner and the priority of execution should be subjected to the RPN ranking. Not all FMEAs follow the same pattern of action plan but the following steps are usually considered

- **Recommended actions –** It consists of all the suggested proceedings that need to be followed to prevent failures. The reasons for failures are multifaceted. Every failure can have several causes; that is why recommended preventive actions are better generated by a cross-functional team

- **Task owner and projected completion date**: The task owner is the person who has been assigned the task of implementing the recommended actions. The projected completion date is also determined to avoid procrastination and enforce accountability

Identify Critical Causes / X's

C Statistical Tools

HYPOTHESIS TESTING

Descriptive & Inferential Statistics

Statistics are often divided into two branches: **Descriptive** and **Inferential** statistics:

Descriptive statistics focus on the collection, analysis, presentation and description of a set of data.

Inferential statistics focus on making decisions about a large set of data, called the population, from a subset of the data, called the sample.

DESCRIPTIVE AND INFERENTIAL STATISTICS

Let's say there are 20 statistics classes at your university, and you've collected the ages of students in one class.

Ages of students in your statistics class: 19, 21, 18, 18, 34, 30, 25, 26, 24, 24, 19, 18, 21, 49, 27.

A descriptive question that could be asked about this data: "What's the most common age of student in your statistics class?" The answer in this case would be 18.

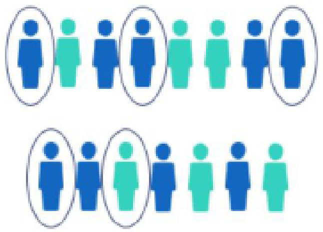
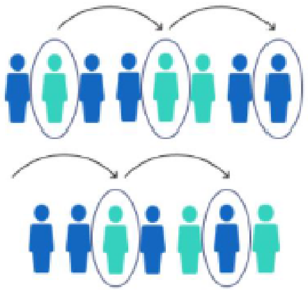
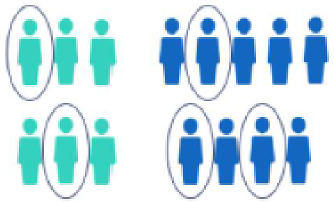
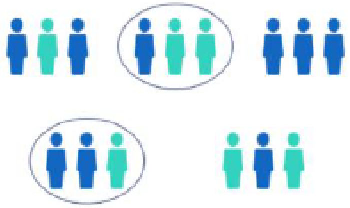
An inferential question that could be asked about this data: "Are the ages of the students in this classroom similar to what you would expect in a normal statistics class at this university?"

DESCRIPTIVE AND INFERENTIAL STATISTICS

S. No	Descriptive Statistics	Inferential Statistics
1	Concerned with the describing the target population	Make inferences from the sample and generalize them to the population.
2	Organize, analyze and present the data in a meaningful manner	Compares, test and predicts future outcomes.
3	Final results are shown in form of charts, tables and Graphs	Final result is the probability scores.
4	Describes the data which is already known	Tries to make conclusions about the population that is beyond the data available.
5	Tools- Measures of central tendency (mean/median/ mode), Spread of data (range, standard deviation etc.)	Tools- hypothesis tests, Analysis of variance etc.

Sampling techniques (Probability)

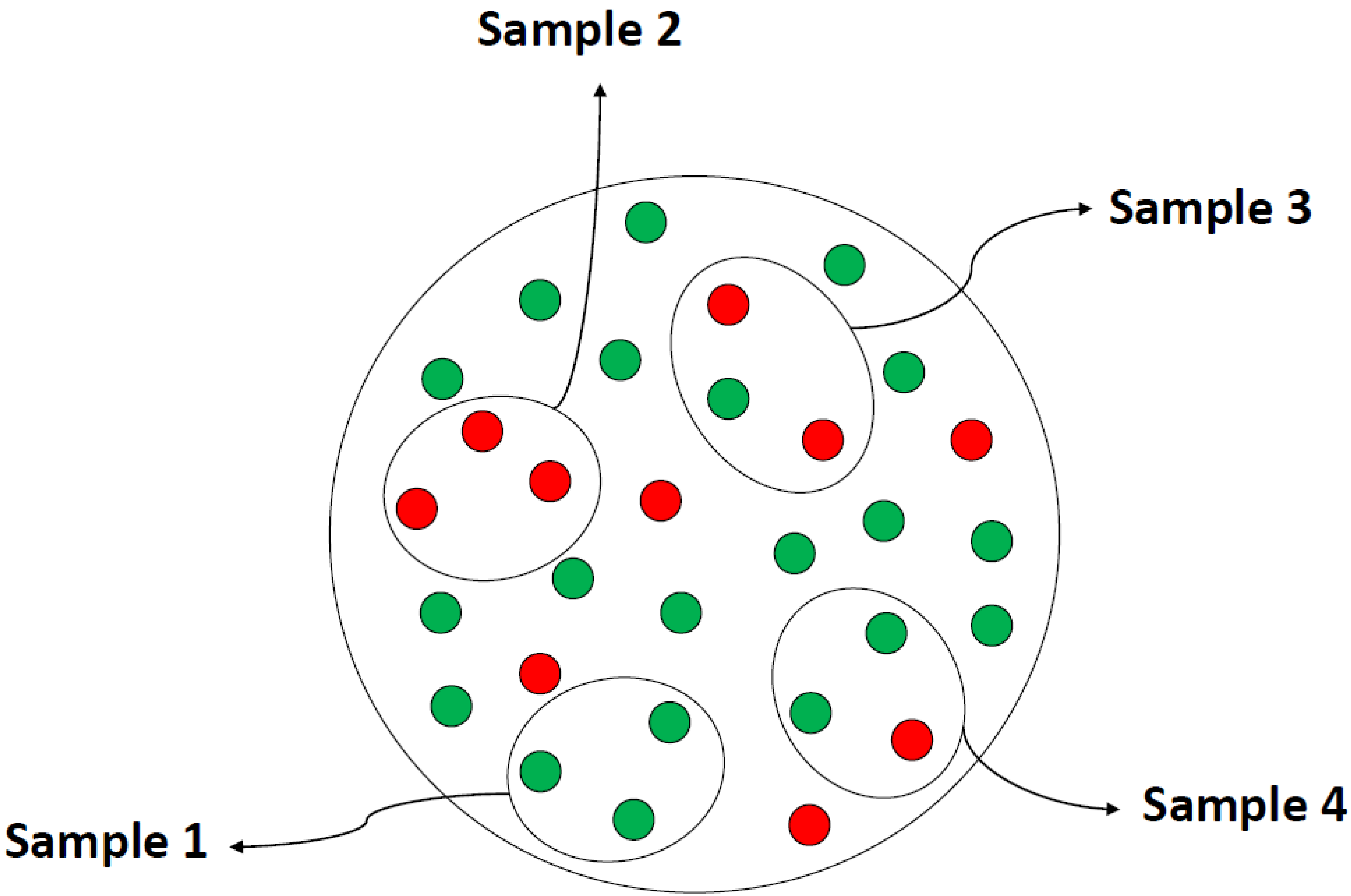
C Statistical Tools

Simple random sampling	Systematic sampling	Stratified sampling	Cluster sampling
Every unit of the population has an equal chance of being selected. Ensures unbiased selection. Generally is a part of other sampling techniques.	Similar to simple random sampling, but instead of randomly generating numbers, units are chosen at regular intervals. Appropriate when data comes in a timely fashion.	Population is divided into subgroups (called strata) based on a relevant characteristic. Random or systematic sampling is used to select a sample from each subgroup. Appropriate when the population has mixed characteristics, and we want every characteristic to be proportionally represented in the sample.	Involves dividing the population into smaller groups known as clusters. Instead of sampling individuals, entire subgroups are randomly selected. This method is good for dealing with large and dispersed (geographically) populations.
			

Hypothesis Testing

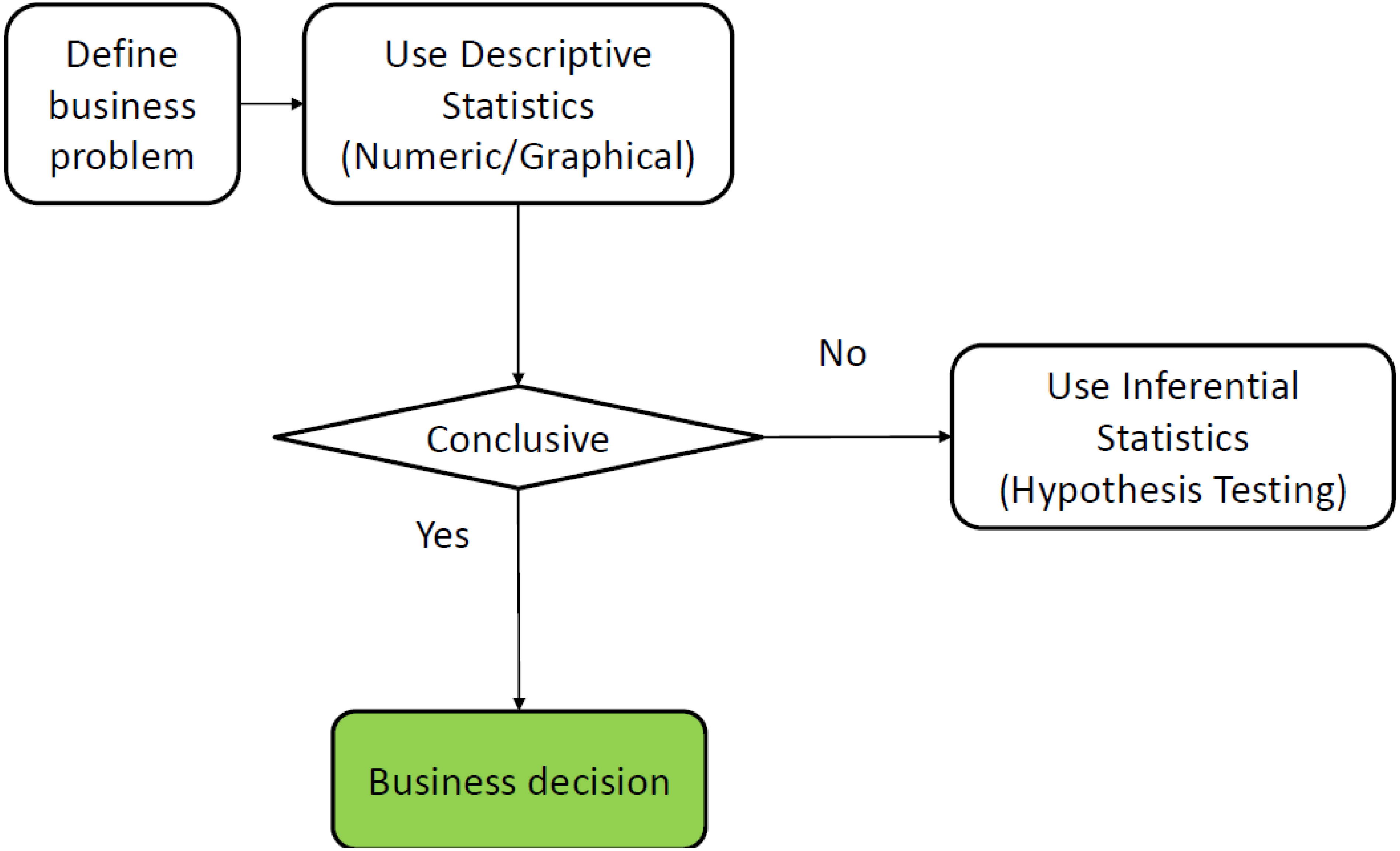
- Hypothesis Testing is a systematic method for testing a claim or hypothesis about a parameter of a population, through a sample. A Hypothesis Test specifies whether to accept or reject this claim depending on the evidence provided by the sample.
- When we estimate the properties of a population from a sample, the sample statistics are unlikely to be exactly equal to the actual population values. The difference between the sample statistic and the population value is the Sampling Error.
- If Sampling Error causes the observed difference, the next time someone performs the sampling again, the results might be different. Hypothesis Testing incorporates estimates of the sampling error to help us make the correct decision.

Sampling Error

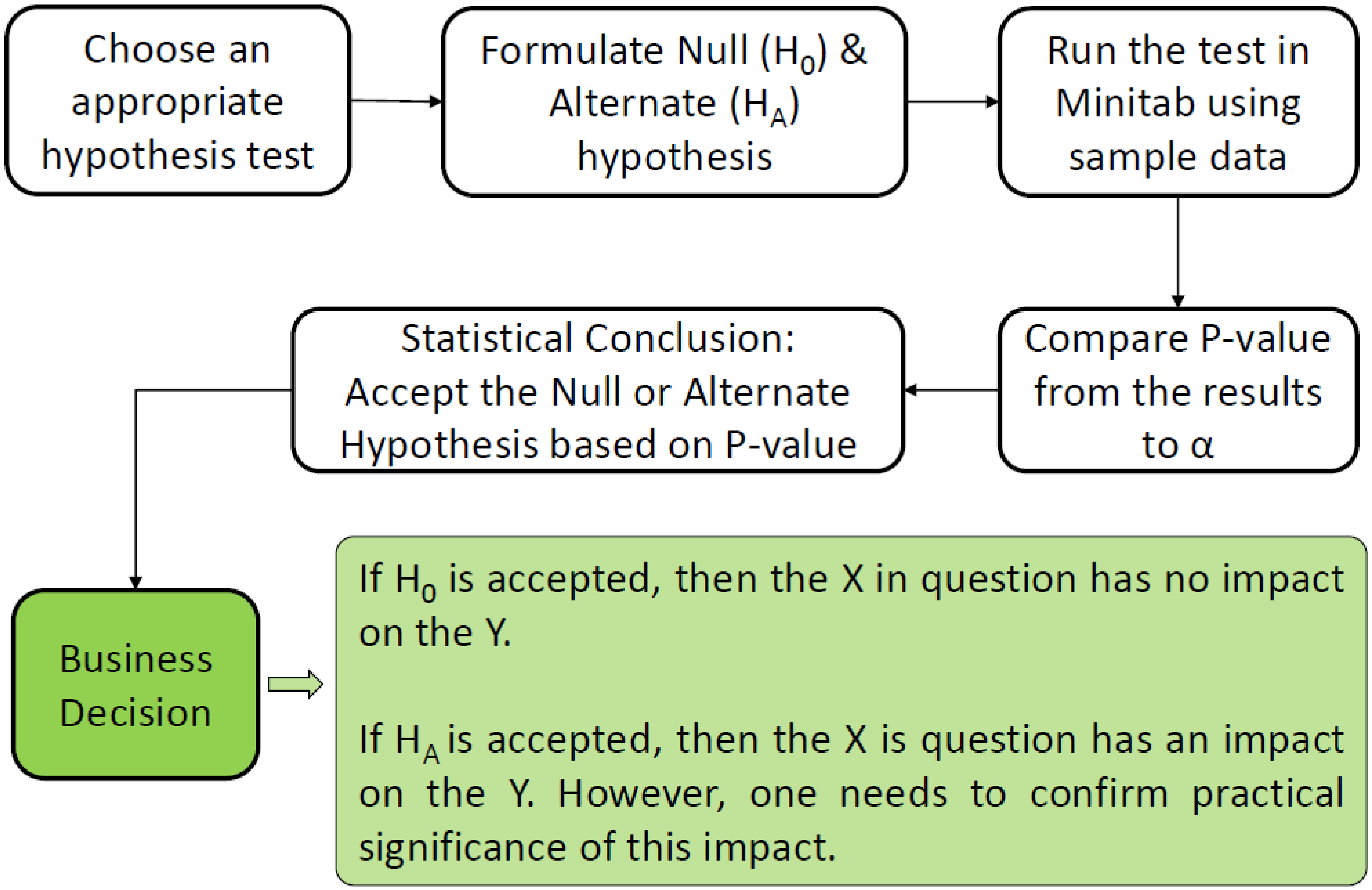


When to use Hypothesis testing?

C Statistical Tools



Hypothesis Testing Flow



Hypothesis Test Selection

C Statistical Tools

Output "Y" is…	Input "X" is …	Appropriate Test	Purpose
Discrete	Discrete in 2 categories	2-proportion test	Compare 2 proportions/percentages
Discrete	Discrete in more than 2 categories	Chi-square test	Compare > 2 proportions/percentages
Continuous	Discrete in 2 categories	2- sample t-test	Compare 2 means (averages)
Continuous	Discrete in more than 2 categories	One-Way ANOVA	Compare > 2 means (averages)
Continuous	Continuous	Scatter Plot, Correlation Linear Regression	Understand relationship between two continuous variables

HYPOTHESIS TESTING

H_0 and H_A Statements

Develop the hypothesis for population and make statistical decisions by determining the acceptance of the hypothesis using sample data.

- ✓ Null Hypothesis (H_0) : The "null hypothesis" assumes that there are no differences or relationships. This is the default assumption of all statistical tests.
- ✓ Alternative Hypothesis (H_A) : The "alternative hypothesis" states that there is a difference or relationship

Errors in Hypothesis Testing

HYPOTHESIS TESTING

	Truth about the population	
Decision based on sample	H_0 is True	H_A is True
Accept H_0	Correct Decision (probability = 1 - α)	**Type II Error** - fail to reject H_0 when it is false (probability = β)
Accept H_A	**Type I Error** - rejecting H_0 when it is true (probability = α)	Correct Decision (probability = 1 - β)

HYPOTHESIS TESTING

Types of Errors

- The **alpha risk or Type 1 Error** (generally called the "Producer's Risk") is the probability that we could be wrong in saying that something is "different." It is an assessment of the likelihood that the observed difference could have occurred by random chance. Alpha is the primary decision-making tool of most statistical tests. It is decided up-front.

- The Confidence level is one minus the alpha level ($1-\alpha$)

- The **beta risk or Type 2 Error** (also called the "Consumer's Risk") is the probability that we could be wrong in saying that two or more things are the same when, in fact, they are different. It is determined from the circumstances of the situation. If alpha is made very small, then beta increases (all else being equal).To minimize beta, while holding alpha constant, requires increased sample sizes

- One minus Beta is the probability of rejecting the null hypothesis when it is false. This is referred to as the Power of the test.

HYPOTHESIS TESTING

P-value

- Decision-making for a Hypothesis Test is done using the **p-value**.
- P-value is the probability that we would obtain the effect observed in our sample, or larger, if the null hypothesis were true. In simpler terms, p-values tell us how strongly our sample data contradict the null.
- **Lower p-values represent stronger evidence against the null.**

> If *p* is less than α, reject H_0 and accept H_A
> If *p* is greater than α, accept H_0 and reject H_A

- Generally, the acceptance level of a Type I error is **0.05.** Thus, any p-value less than 0.05 means we reject the Null Hypothesis.
- ***P value applies for ALL Hypothesis tests!***

HYPOTHESIS TESTING

The length of 25 samples of a fabric are taken at random to get the following data. Mean and standard deviation from the historic 1-year study are 150 and 4 respectively. Test if the current mean is greater than the historic mean. Assume α to be 0.05.

151.2	160.3	147.5	149.2	159.2	155.6	148.1	154.8	156.8
167.5	164.5	147.9	154.1	159.4	147.5	148.2	154.2	148.1
163.1	155.2	154.6	155.4	158.2	157.6	158.4		

HYPOTHESIS TESTING

C Statistical Tools

151.2	
160.3	
147.5	
149.2	
159.2	
155.6	
148.1	
154.8	
156.8	
167.5	
164.5	
147.9	
154.1	
159.4	
147.5	
148.2	
154.2	
148.1	
163.1	
155.2	
154.6	
155.4	
158.2	
157.6	
158.4	
MEAN	155.064
MEDIAN	155.2

sample size(n)	25
σ	4
μ(population/historic mean)	150
x̄ (sample mean)	155.064
Z	6.33
P	0.00000
H0:μ=150	
H1:μ>150	

WE ACCEPT ALTERNATIVE HYPOTHESIS

C Statistical Tools

HYPOTHESIS TESTING

	F	G	H	I
14				
15				
16	sample size(n)	25		
17	σ	4		
18	μ(population/historic mean)	150		
19	x̄ (sample mean)	155.064		
20				
21	Z	=(G19-G18)/(G17/SQRT(G16))	6.33	
22	P	0.00000		

C Statistical Tools

HYPOTHESIS TESTING

	F	G
14		
15		
16	sample size(n)	25
17	σ	4
18	μ(population/historic mean)	150
19	x̄ (sample mean)	155.064
20		
21	Z	6.33
22	P	=1-NORMSDIST(G21)

0=1-1

HYPOTHESIS TESTING

151.2	
160.3	
147.5	
149.2	
159.2	
155.6	
148.1	
154.8	
156.8	
167.5	
164.5	
147.9	
154.1	
159.4	
147.5	
148.2	
154.2	
148.1	
163.1	
155.2	
154.6	
155.4	
158.2	
157.6	
158.4	
MEAN	155.064
MEDIAN	155.2

sample size(n)	25
σ	4
μ(population/historic mean)	150
x̄ (sample mean)	155.064
Z	6.33
P	0.00000
H0:μ=150	
H1:μ>150	

Test

Null hypothesis H_0: μ = 150

Alternative hypothesis H_1: μ > 150

Z-Value	P-Value
6.33	0.000

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

State Results

Decision Rule: If Z is less than -1.96, or greater than 1.96, reject the null hypothesis.

WE ACCEPT ALTERNATIVE HYPOTHESIS

TO BE CONTINUED IN NEXT MODULE