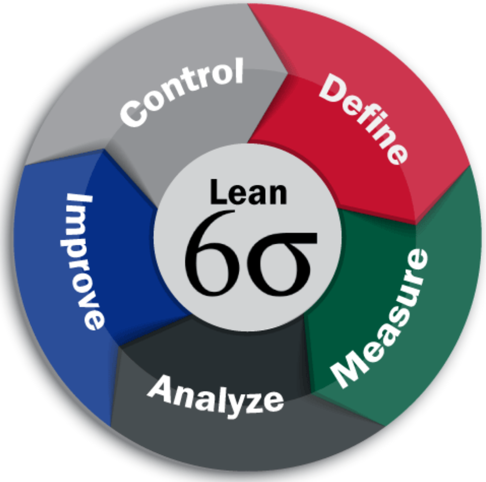
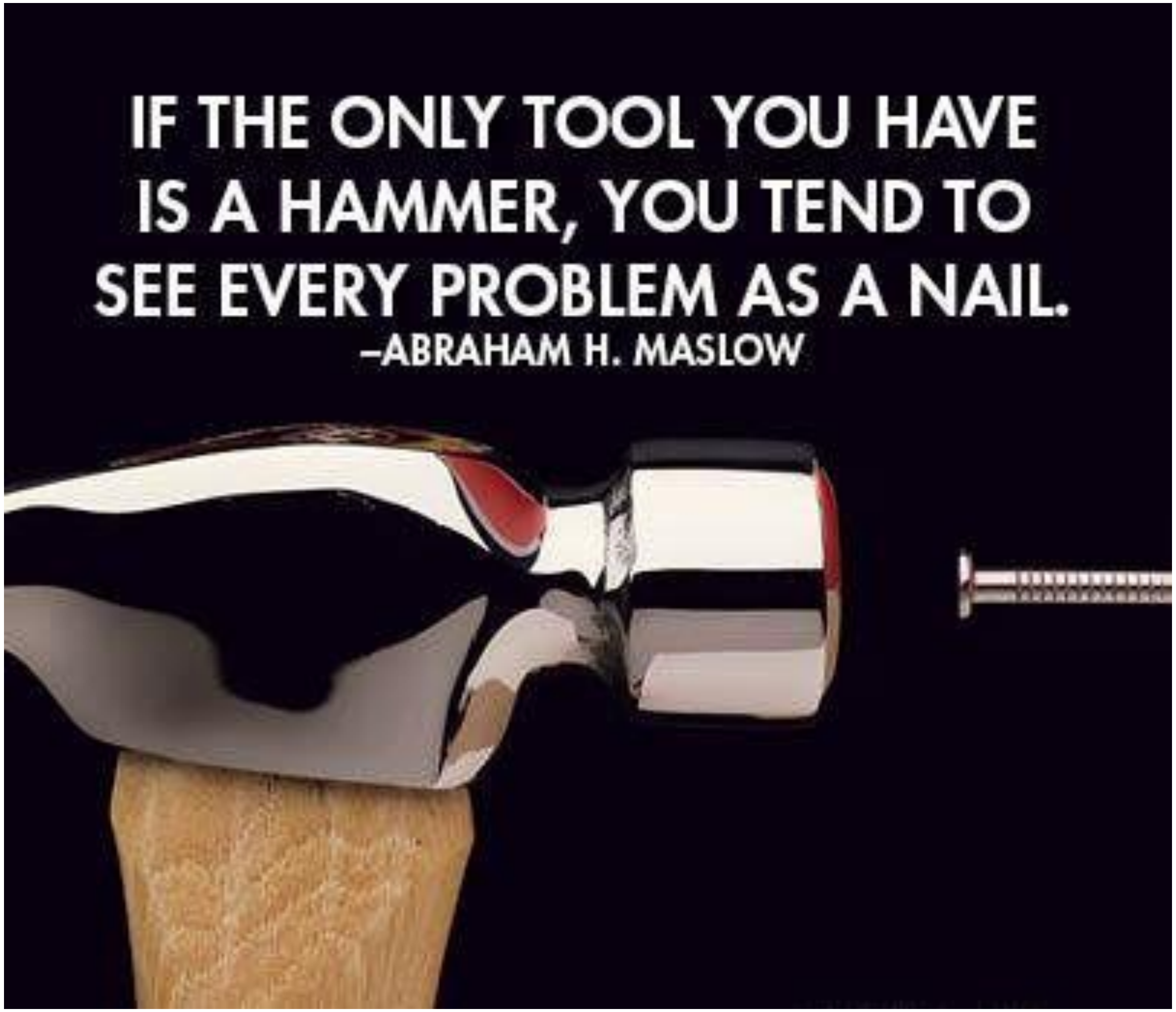


LEAN SIX SIGMA ANALYZE PHASE

"If you torture the data long enough, it will confess." - Ronald H. Coase

Coverage:

- ANALYZE PHASE TOOLS
  - 7 QC Tools (Histogram, Cause-and-Effect Diagrams / Fishbone Diagram, Pareto charts, Control Charts, Scatter diagrams, Stratification, Check Sheet) and PFMEA
  - 5 Why Analysis, Tree Diagram, Affinity clustering
  - Data visualization - Box Plot , Multi Vari charts Trends and comparison charts
  - Hypothesis testing - Parametric test (Continuous Data , Discrete Data) Alpha & Beta errors

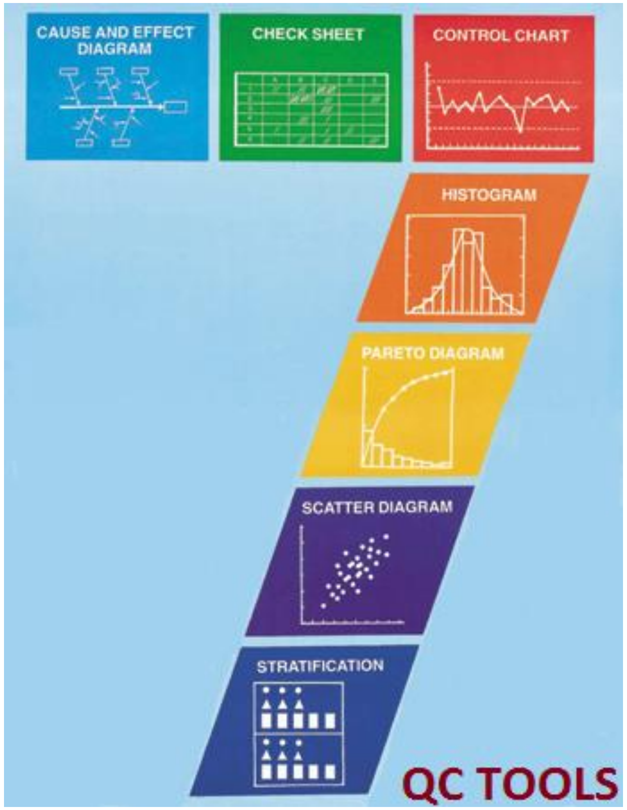

IF THE ONLY TOOL YOU HAVE
IS A HAMMER, YOU TEND TO
SEE EVERY PROBLEM AS A NAIL.
–ABRAHAM H. MASLOW

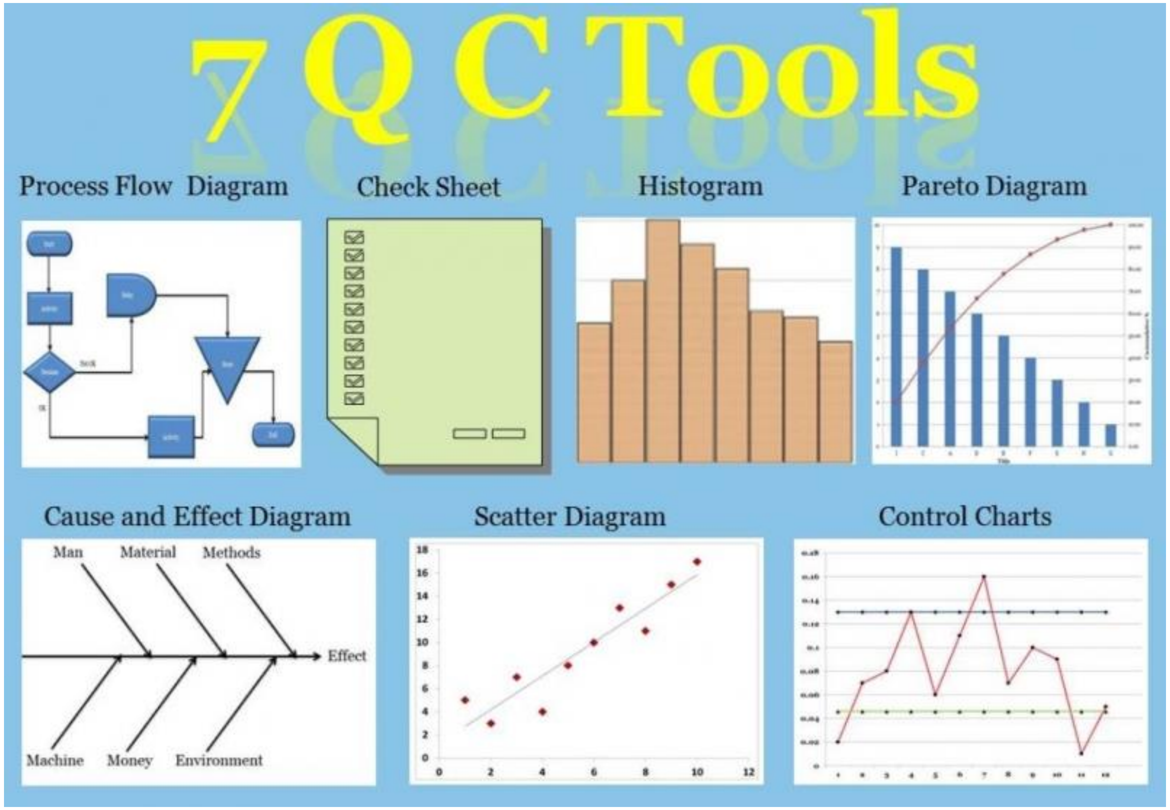
Analyze Phase

The Analyze phase of DMAIC project helps project teams to identify “Vital X” form the list of “Potential Xs”.

In this phase project teams use a variety of Process door and Data door analysis tools to validate the root cause of the problem

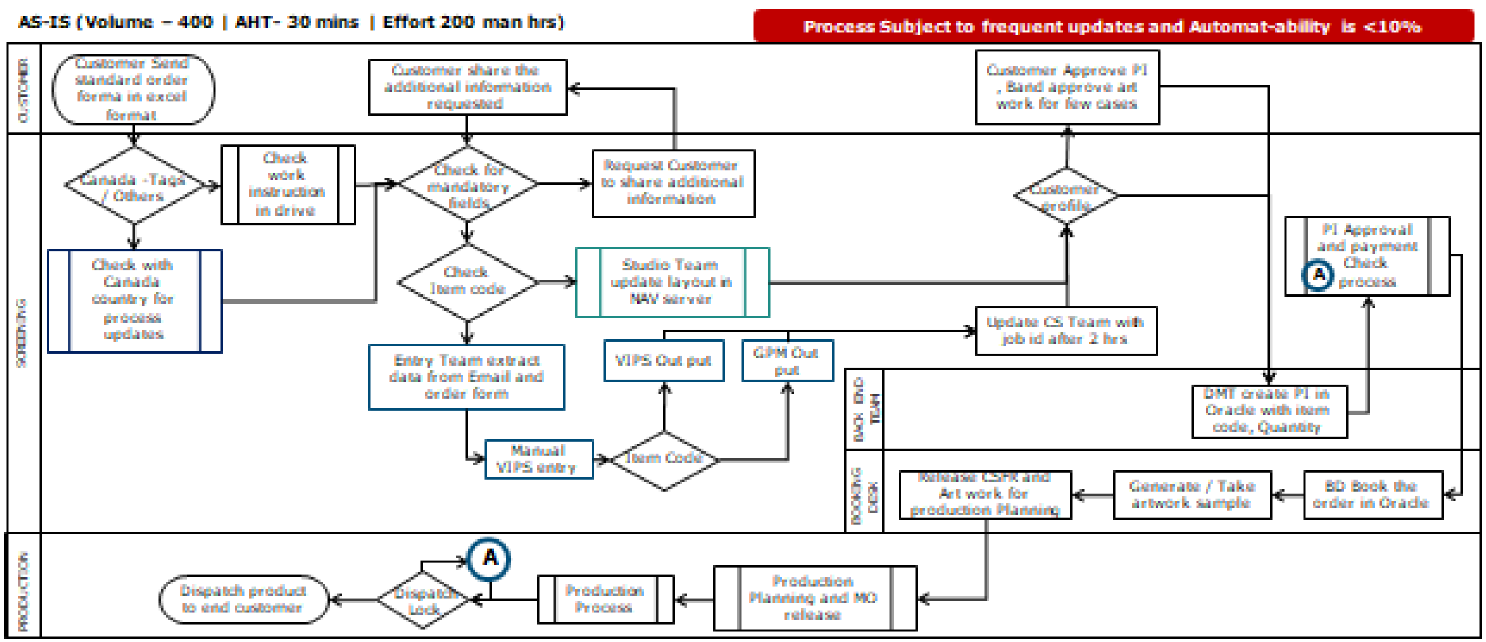
7 QC Tools


CAUSE AND EFFECT DIAGRAM
CHECK SHEET
CONTROL CHART
HISTOGRAM
PARETO DIAGRAM
SCATTER DIAGRAM
STRATIFICATION
QC TOOLS


7 Q C Tools
Process Flow Diagram
Check Sheet
Histogram
Pareto Diagram
Cause and Effect Diagram
Man
Material
Methods
Effect
Machine
Money
Environment
Scatter Diagram
Control Charts

7 QC Tools – Process Flow

Process Flow is a graphical representation of a business process through a flowchart. It's used as a means of getting a top-down understanding of how a process works, what steps it consists of, what events change outcomes, and so on



7 QC Tools – Process Flow

Stage 1

- Request for a documented process (SOP) form the process owner
- Study the documented process
- Create a process flow based on your understanding
- Check with the process owner if your process flow is inline with the document process

Stage 2

- Spend a "Day in the Life of (DILO)" the processor and work-shadow the processor
- Pick up random transaction and travel to different departments and track how the transaction is processed

Stage 3

- Highlight the deviations observed in the shop floor and the documented process to the process owner

Stage 4

- Request the process owner to confirm the final version with written approval

Stage 5

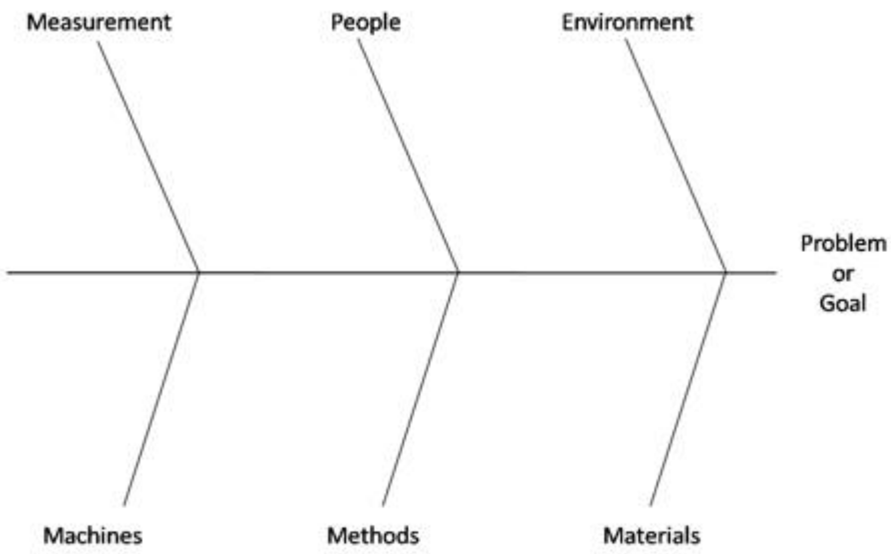
- Document the final version of process flow in project PPT

7 QC Tools – Check Sheet

Problem	Material						Total
	X			Y			
	Shift			Shift			
	1	2	3	1	2	3	
A	18	21	22	23	30	28	142
B	7	6	5	8	9	8	43
C	12	11	24	17	15	17	96
D	14	13	8	5	2	4	46
Total	51	51	59	53	56	57	
	161			166			327

Check sheet one of the 7QC tools it is a form or data collection template used to collect data in real time at the location where the data is generated. The data it captures can be quantitative or qualitative

7 QC Tools – Cause & Effect Diagram



Kaoru Ishikawa

A Cause and Effect Diagram is a graphical tool for displaying a list of causes associated with a specific effect. It is also known as a fishbone diagram or an Ishikawa diagram (created by Dr. Kaoru Ishikawa, an influential quality management innovator). The graph organizes a list of potential causes into categories.

7 QC Tools – Cause & Effect Diagram

What is Cause And What is effect ?

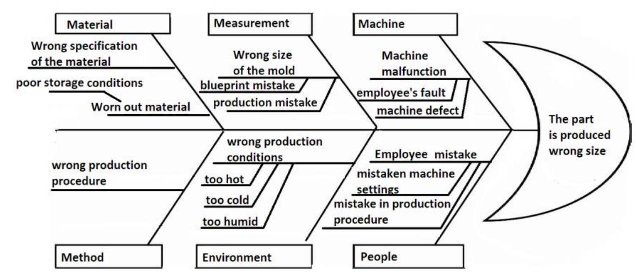
Effect is the problem or the outcome and case is the principle reason for the outcome

Sequence of events

Brainstorming – Clustering – Cause & Effect Diagram

Brainstorming

Clustering



Fishbone diagram

7 QC Tools – Cause & Effect Diagram - Brainstorming

Brainstorming

To brainstorm is to think about and try to come up with Potential cause to a problem or ideas or solutions to a problem.

Brainstorming combines a relaxed, informal approach to problem solving with lateral thinking. It encourages people to come up with thoughts and ideas that can, at first, seem a bit crazy

Brainstorming

Types of Brainstorming

•Analogy

•Brain Writing

•Try Storming

•Anti-Storming

•Step ladder approach

•Questioning Orthodoxy

Dos and Don’ts of brainstorming

•Ambience & atmosphere during discussion must be perfect

•Focus on the theme.

•Give everybody the opportunity to speak.

•No criticism of ideas as they are expressed.

•Do not get into action with only one idea.

•Do not get into details.

•Do not evaluate an idea.

•Do Not Paraphrase

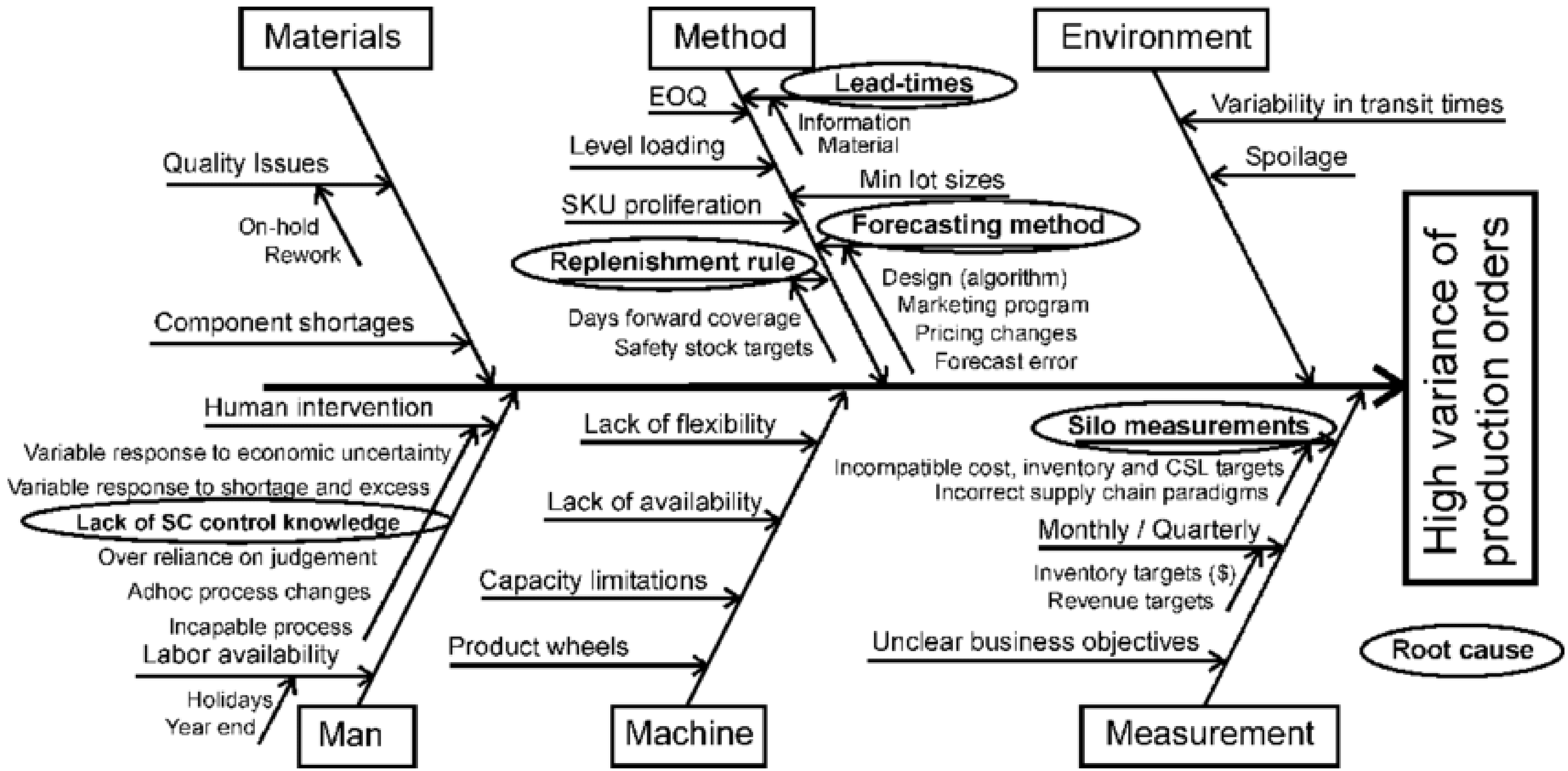
7 QC Tools – Cause & Effect Diagram –clustering

Clustering is a method which can help you gather large amounts of data and organize them into groups or themes based on their relationships. The affinity process is great for grouping data gathered during research or ideas generated during Brainstorming

Brainstorming Outcome

Clustering

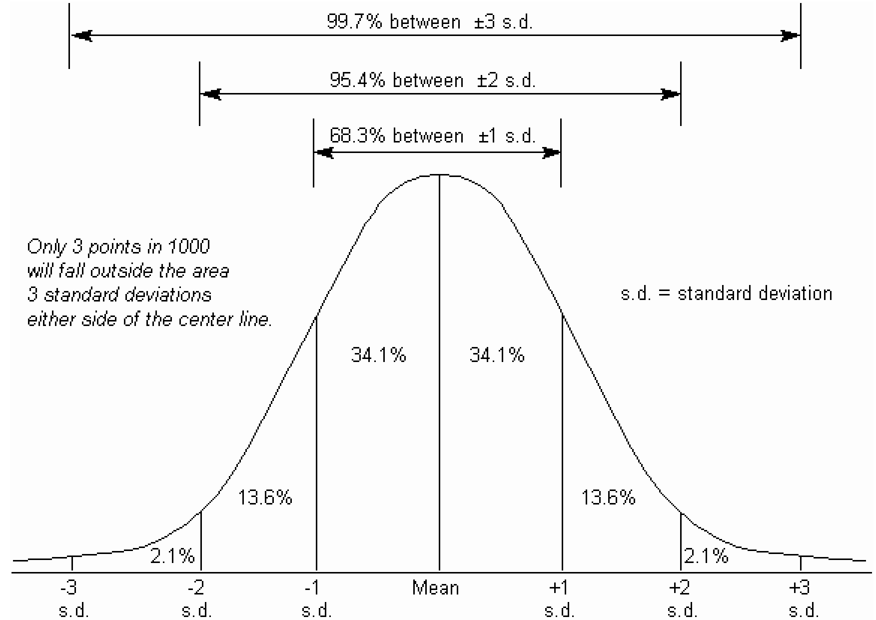
7 QC Tools – Cause & Effect Diagram



7 QC Tools – Histogram

Histogram is a graphical representation of the distribution of numerical data

Purpose : The shape, mean & spread of the histogram gives additional information about the process and guides the managers to take the right decision to improve the process.



How to construct histogram manually

Step 1 : Calculate the range (Maximum Value – Minimum Value)

Step 2 : Create class intervals (about 8 or 10) by determining appropriate class intervals

Step 3 : Create frequency distribution table

Step 4 : Draw histogram based on the frequency distribution table

Class Width = 2

Classes	Frequency	Total
34-35	𝍸 |||	8
36-37	𝍸 𝍸 𝍸 𝍸𝍸	25
38-39	𝍸 𝍸 𝍸 𝍸𝍸 𝍸		31
40-41	𝍸 𝍸 𝍸 𝍸𝍸 𝍸 𝍸	35
42-43	𝍸 𝍸 𝍸 𝍸	20
44-45	𝍸	5
46-47			1

7 QC Tools – Pareto chart

Vilfredo Pareto was an Italian Economist in the 19th Century In 1906, he made the famous observation that twenty percent of the population owned eighty percent of the property in Italy.

This is famously referred as “80-20 rule” or “Vital few trivial many” and graphical represented as “Pareto Chart”

The tool was popularized by Joseph M Juran and Kaoru Ishikawa.

Vilfredo Pareto

How to construct Pareto chart

Step 1 : Root cause identified during fishbone diagram can be converted as defect categories

Step 2 : Collect the data to identify the frequency of each defect category

Step 3 : Calculate the contribution of each category to the total number of defects

Step 4 : Calculate Cumulative Contribution

Step 5 : Plot the number of defect counts in primary access (As bar-graph)

Step 6 : Plot the cumulative contribution in Secondary access (As Line-graph)

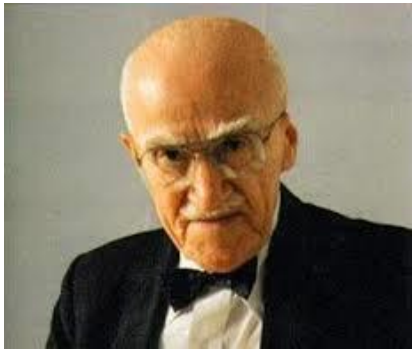
Minitab exercise

Excel Analysis

7 QC Tools – Stratification

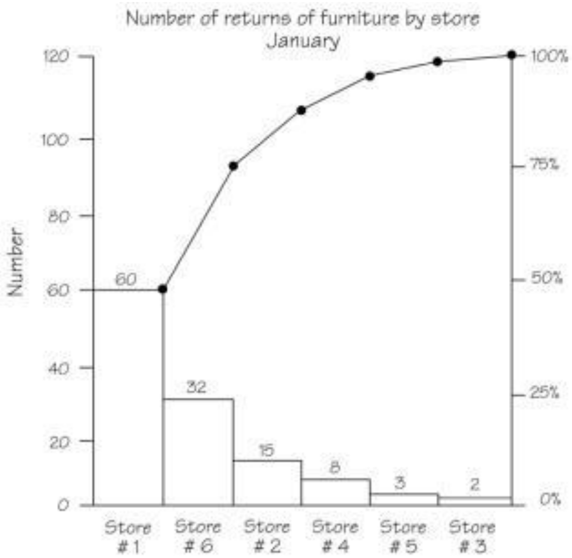
Stratification is to classify or group data with matching characteristics in groups or strata. It serves to facilitate the work before using other tools such as histograms or scatter diagrams

Joseph M Juran in his famous book "Quality Hand Book" Stratification is the separation of data into categories. It is used to identify which categories contribute to the problem being solved and which categories are worthy of further investigation

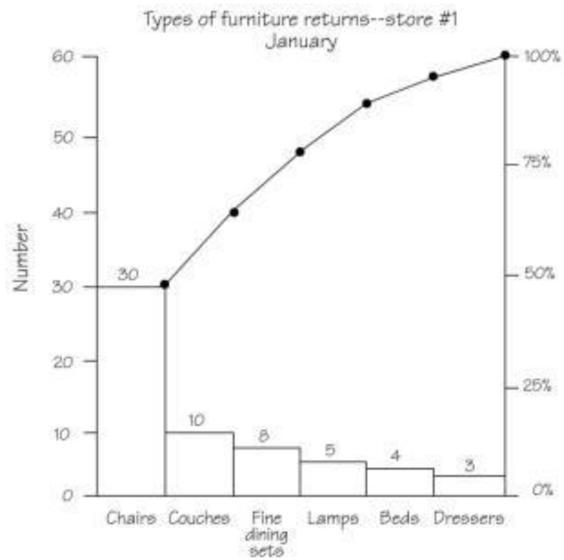


Joseph M. Juran

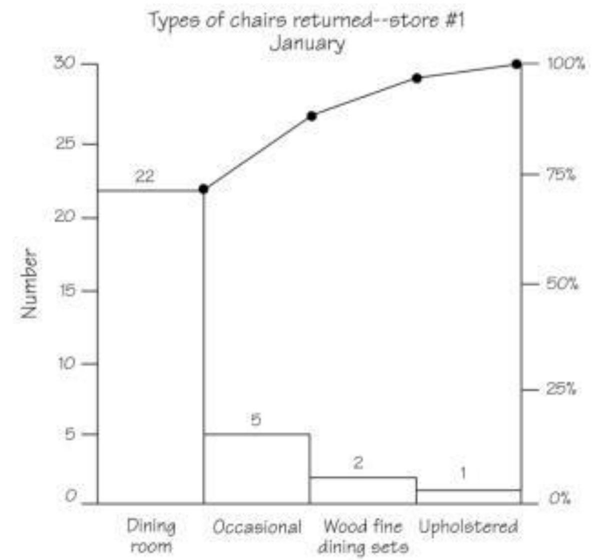
7 QC Tools – Stratification - Subdivisions



Level 1



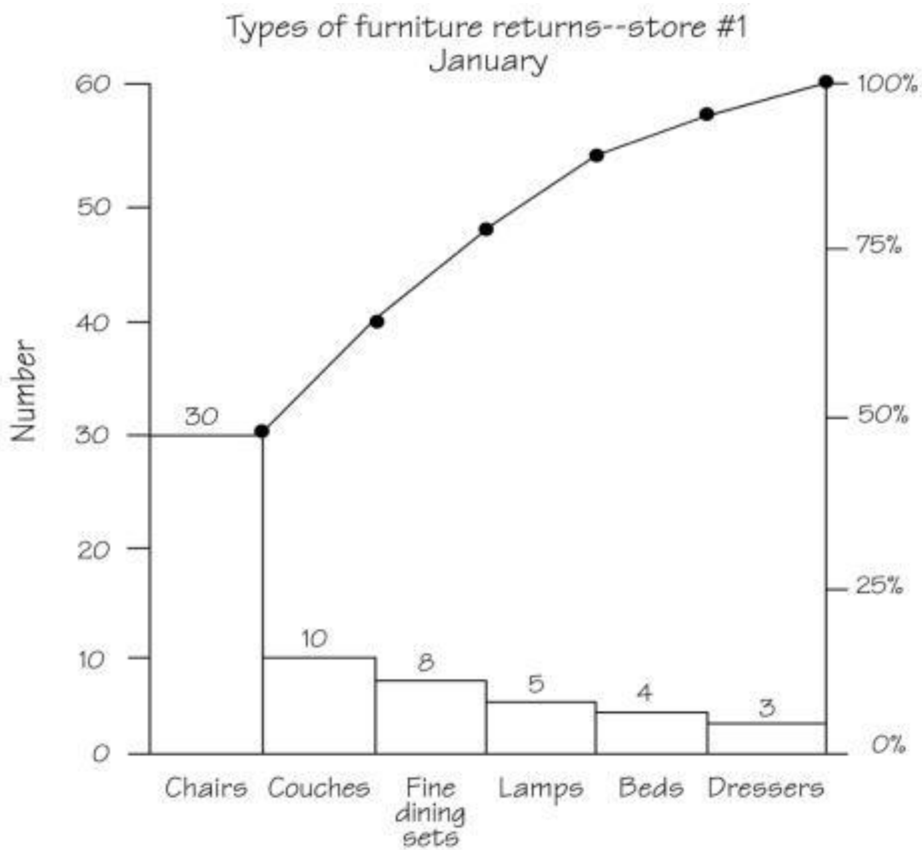
Level 2

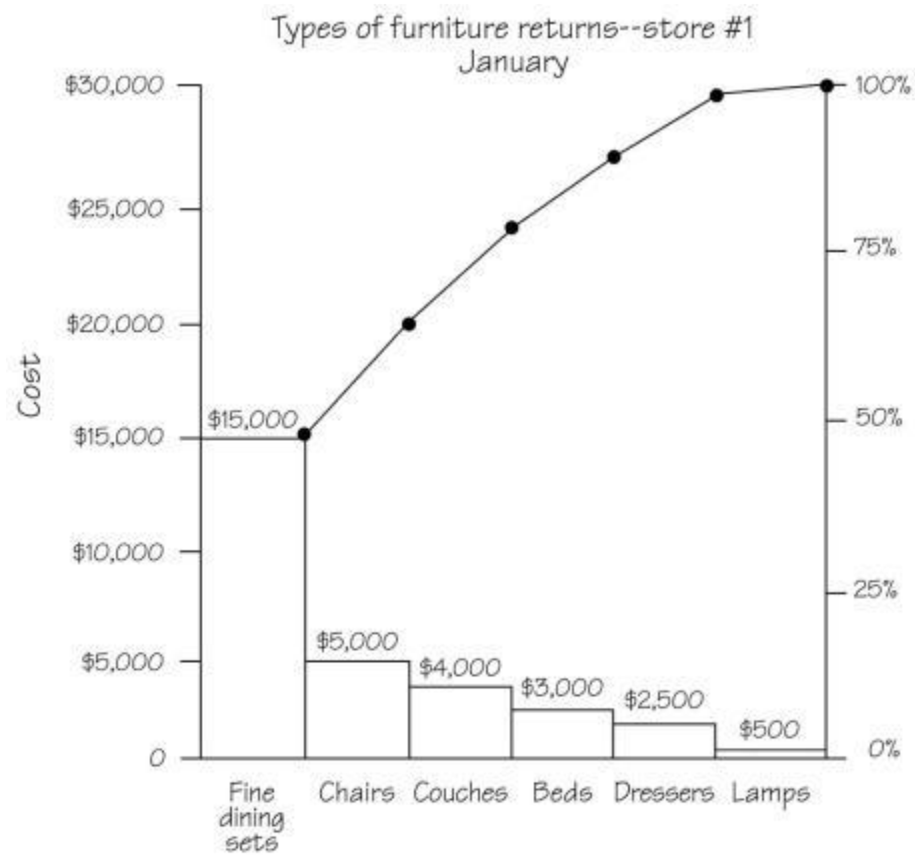


Level 3

Multiple drill down of the same data

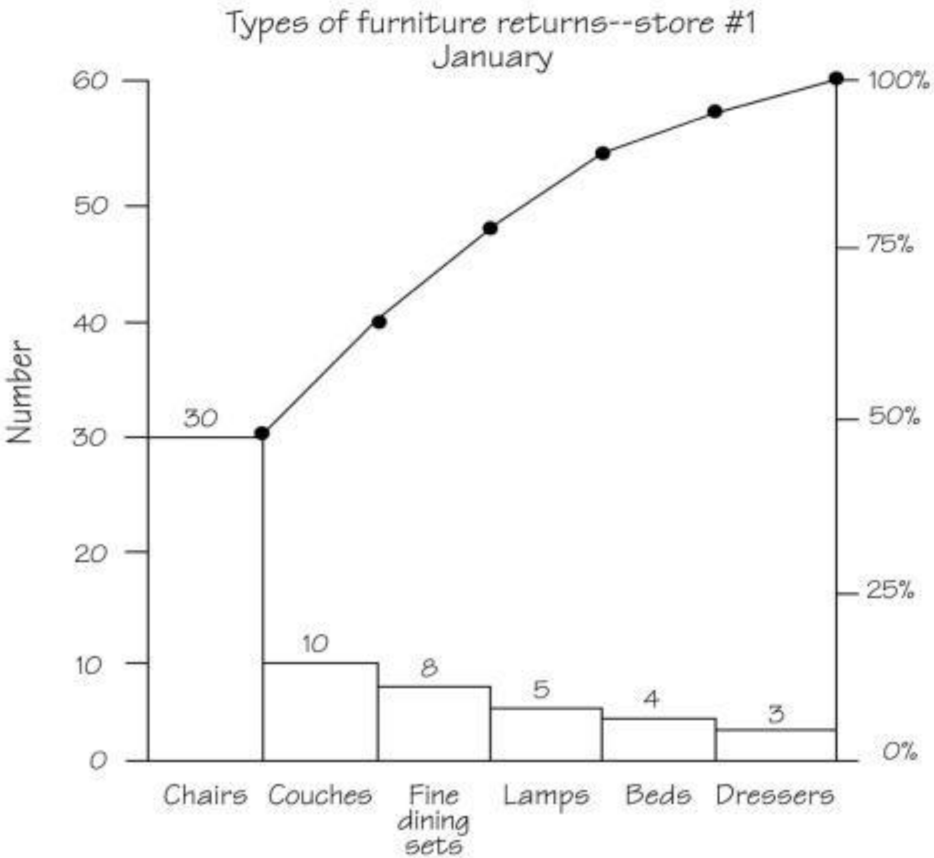
7 QC Tools – Stratification - Multi-perspective analysis

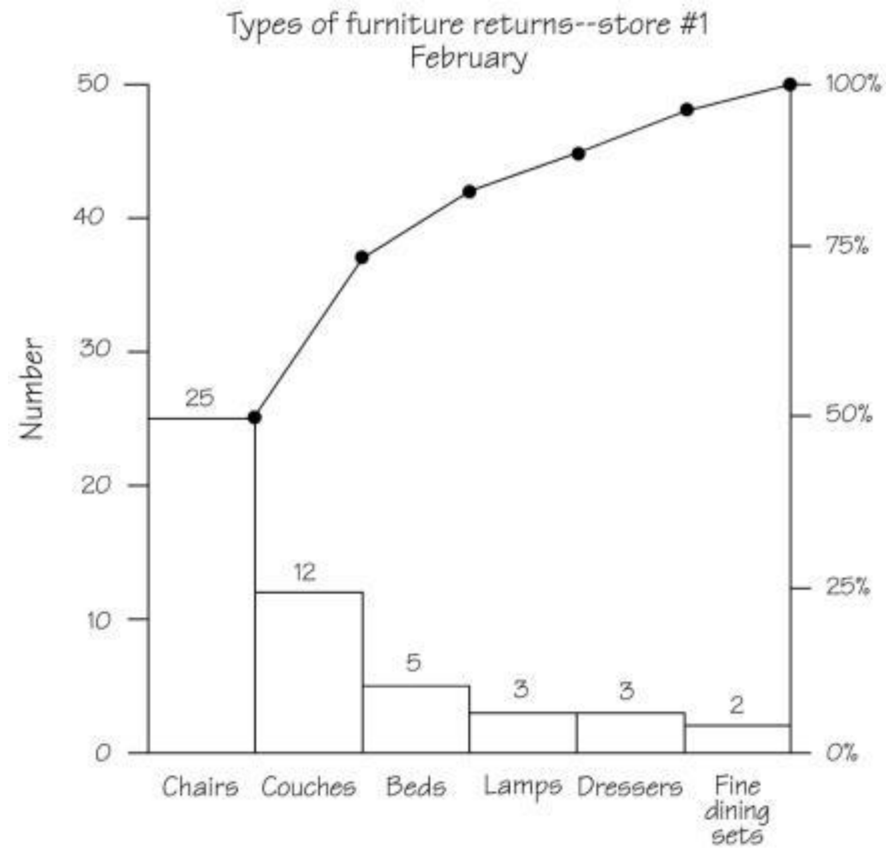




Same return data analyzed in count and $value

7 QC Tools – Stratification - Repeat analysis

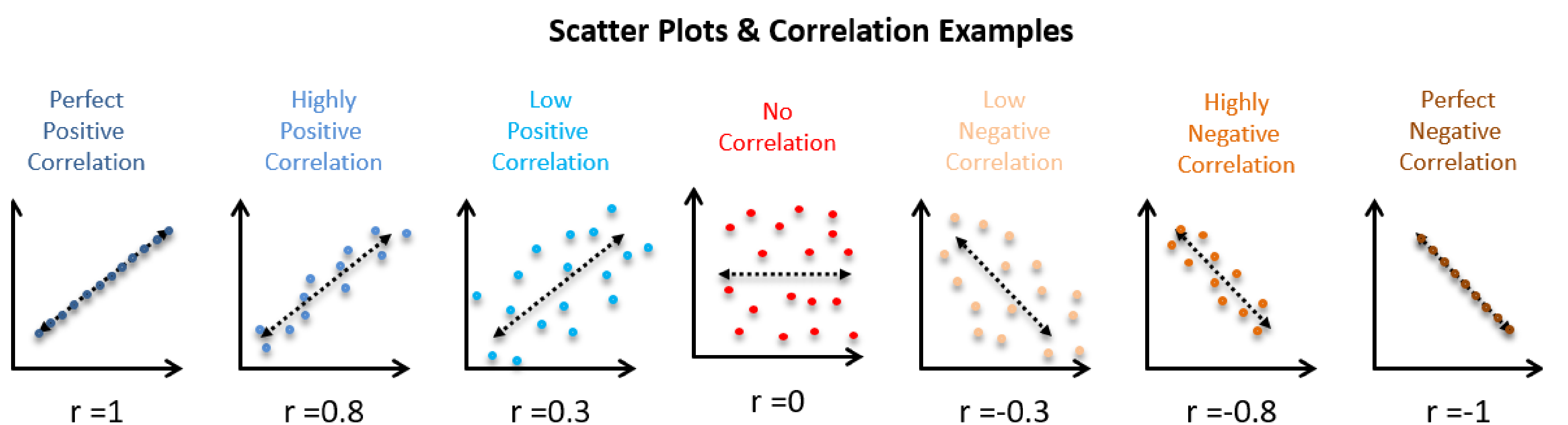


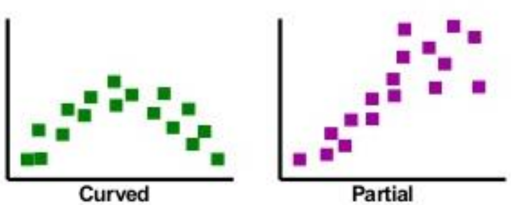


Return data analyzed for two consecutive timeperiod

7 QC Tools – Scatter Diagram

Correlation Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. The strength of correlation is statistically represented “correlation coefficient” and graphically represented by “Scatter Diagram”





Correlation Coefficient level of significance

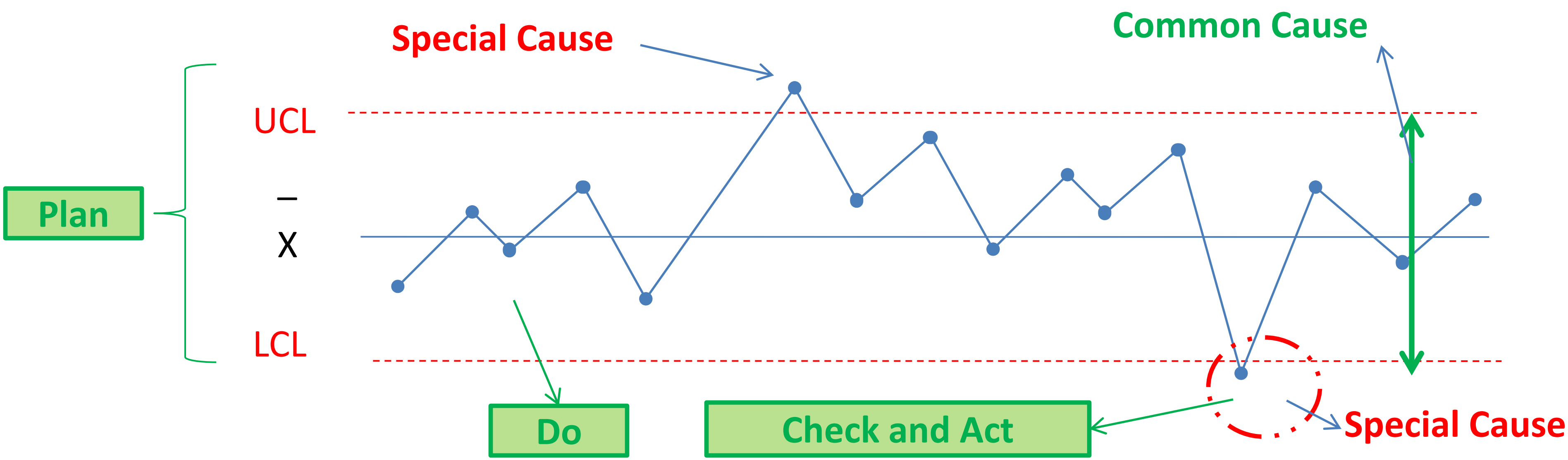
- 0.8 to 1 (Strong Positive correlation)
- -0.8 to -1 (Strong Negative correlation)
- Between 0.7 to 0.5 and -0.7 to -0.5 (week correlation)
- Between 0.5 t0 -0.5 (No correlation)

7 QC Tools – Control Chart

Statistical process control (SPC) is a method of quality control which employs statistical methods to monitor and control a process. Control Chart is one of the important SPC tool

Control Chart also known as Shewhart charts or process-behavior charts, are a statistical process control tool used to distinguish between common cause variation and special cause variation

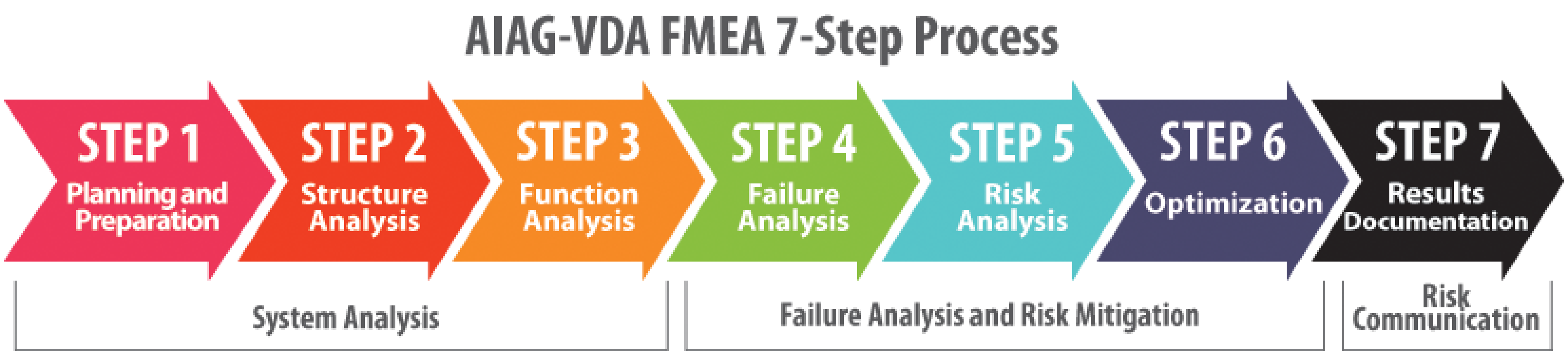
Walter A. Shewhart



Failure Mode Effects Analysis - FMEA

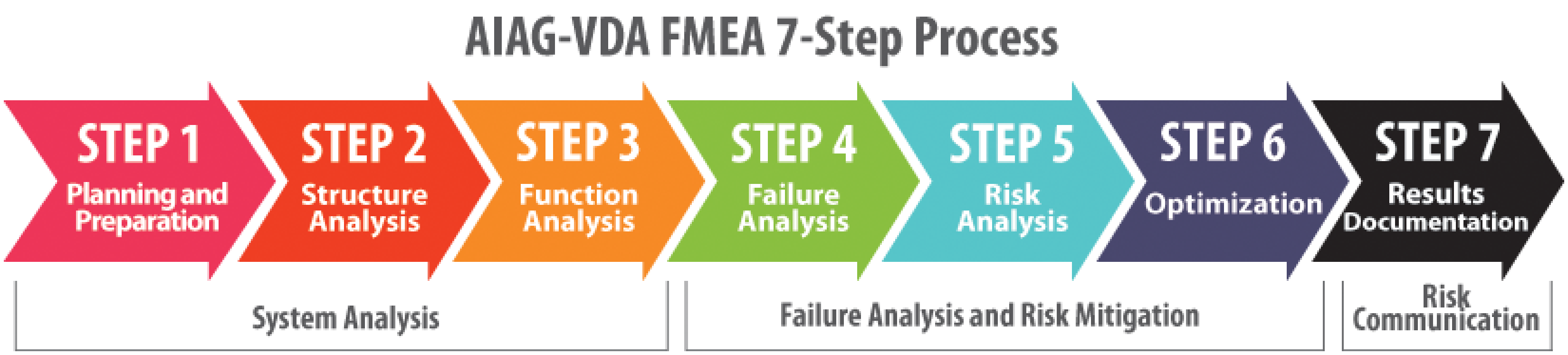
- FMEA is a structured approach to Identifying the areas and ways in which a process or system can fail (failure mode)
- Estimating risk associated with specific causes
- Identifying and prioritizing the actions that should be taken to reduce those risks
- Evaluating and documenting proposed process plans or current control plans
- FMEA is used to reduce risk, and therefore unintended consequences, in the implementation.

PFMEA Process FMEA : A Process Failure Mode Effects Analysis (PFMEA) is a structured analytical tool used by an organization, business unit, or cross-functional team to identify and evaluate the potential failures of a process

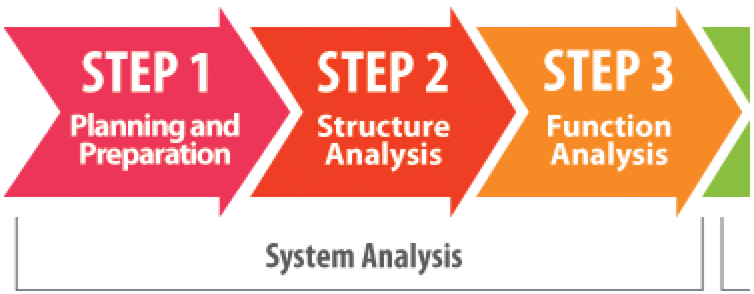


Failure Mode Effects Analysis - FMEA

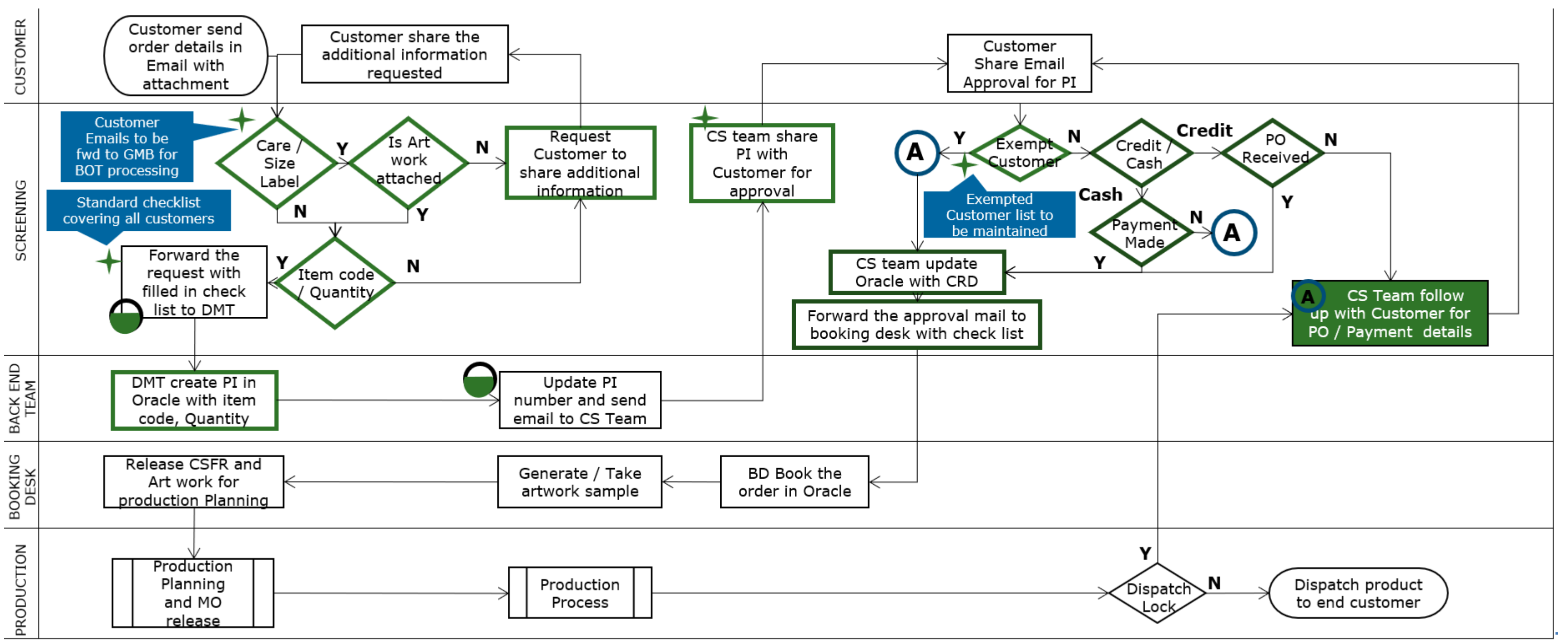
PFMEA Process FMEA : A Process Failure Mode Effects Analysis (PFMEA) is a structured analytical tool used by an organization, business unit, or cross-functional team to identify and evaluate the potential failures of a process



Failure Mode Effects Analysis - FMEA



- Identify all stake holders
- Map the process (Material flow as well as information flow) at a most granular level



Failure Mode Effects Analysis - FMEA

•For each process step determine the ways in which the input to the step or the function of the step can go wrong (failure mode – People , Process , Technology).

•For each failure mode, determine effects - Select a severity rating for each effect (1-10 scale)

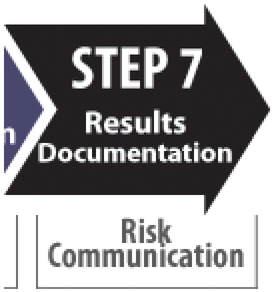
•Identify potential causes of each failure mode - Select an occurrence rating for each cause (1-10 scale)

•List current monitoring and controls for each cause -Select a detection rating for each cause (1-10 scale)

•Multiple these three ratings to calculate Risk Priority Number (S * O * D = RPN)

•Identify the process steps with high RPN (>100) and take required actions to bring down the RPN

Failure Mode Effects Analysis - FMEA

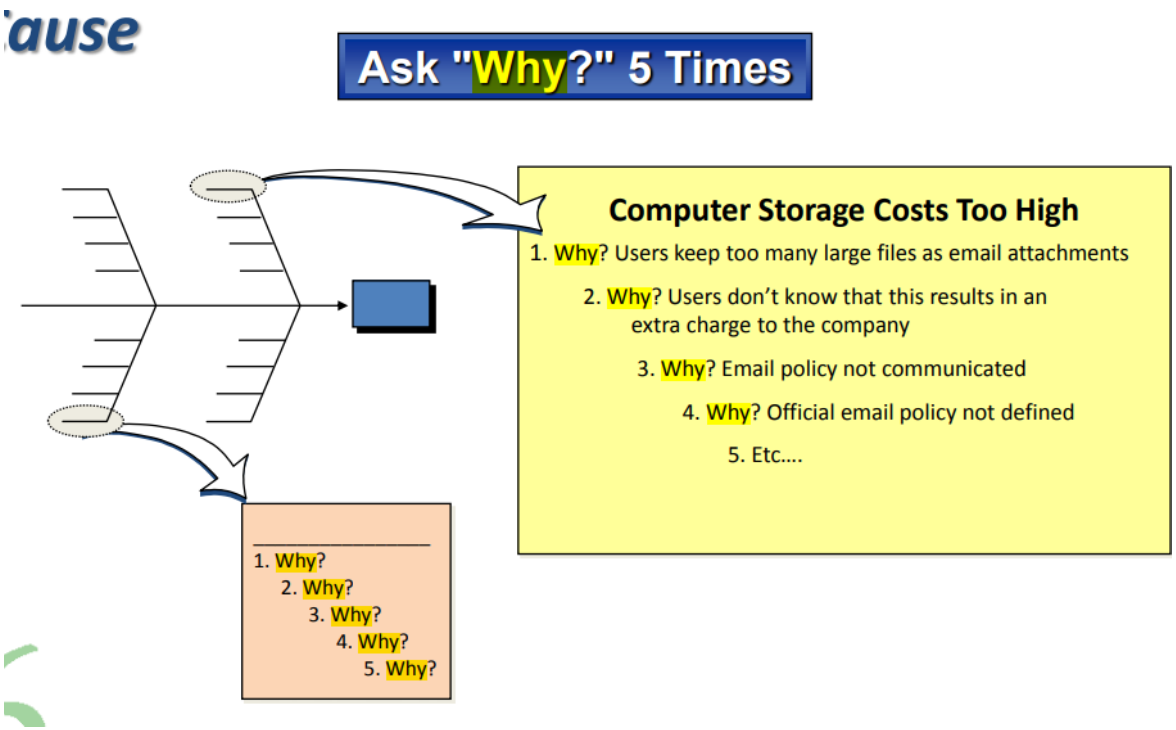


- FMEA should be a living document , Keep updating the actions taken and subsequent reduction in RPA

Assume that the Severity number cannot be reduced. Indicate the order of importance that you would assign as far as addressing
these processes so as to reduce overall risk.

Item	Severity	Occurrence	Detection	RPN
a	8	10	2	160
b	10	8	2	160
c	8	2	10	160
d	10	2	8	160

5 why Analysis



The 5 Whys technique is a simple and effective tool for solving problems. Its primary goal is to find the exact reason that causes a given problem by asking a sequence of "Why" questions.

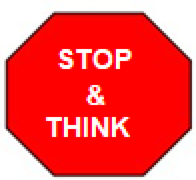
5 why Analysis – What’s wrong with this

The bread on your table is mouldy → Throw it out & get a new piece

The waiter“s focus was a one-time solution,
But the restaurant manager thinks this is a frequent problem.
Let“s use 5 Whys to get to the root cause of mouldy bread…

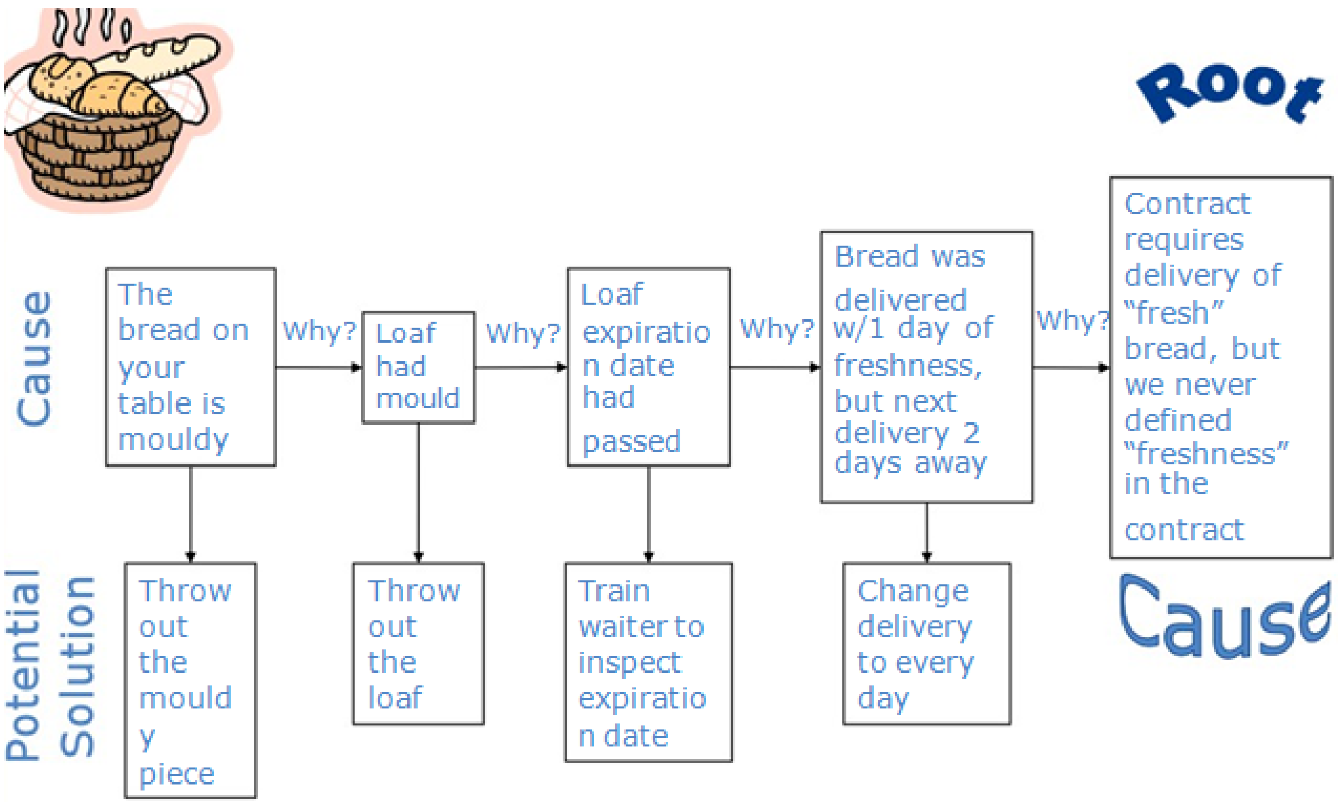
The bread on your table is mouldy —Why?→ Waiter didn“t see the mould —Why?→ Poor training / Old glasses / Poor lighting

Potential solutions:
- Re-do training
- Vision plan
- Inspection station



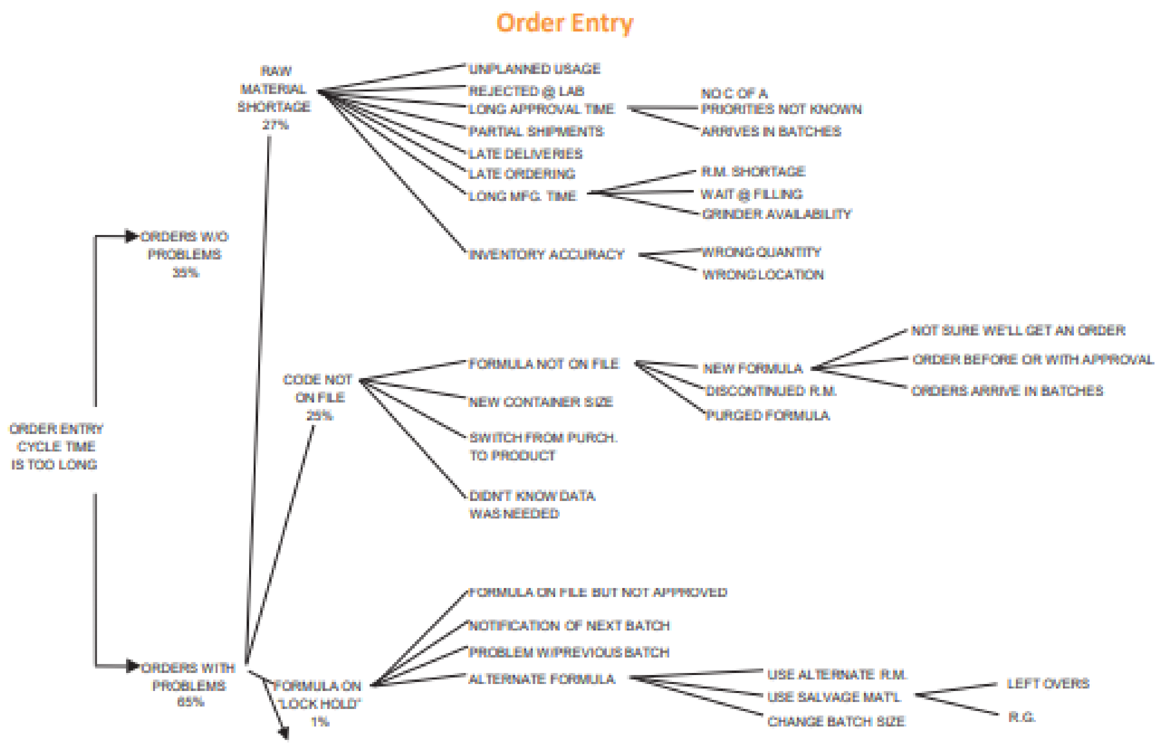
Are these types of solutions preventing the mould from being there in the first place?
Let“s try again…

5 why Analysis – How it should be done

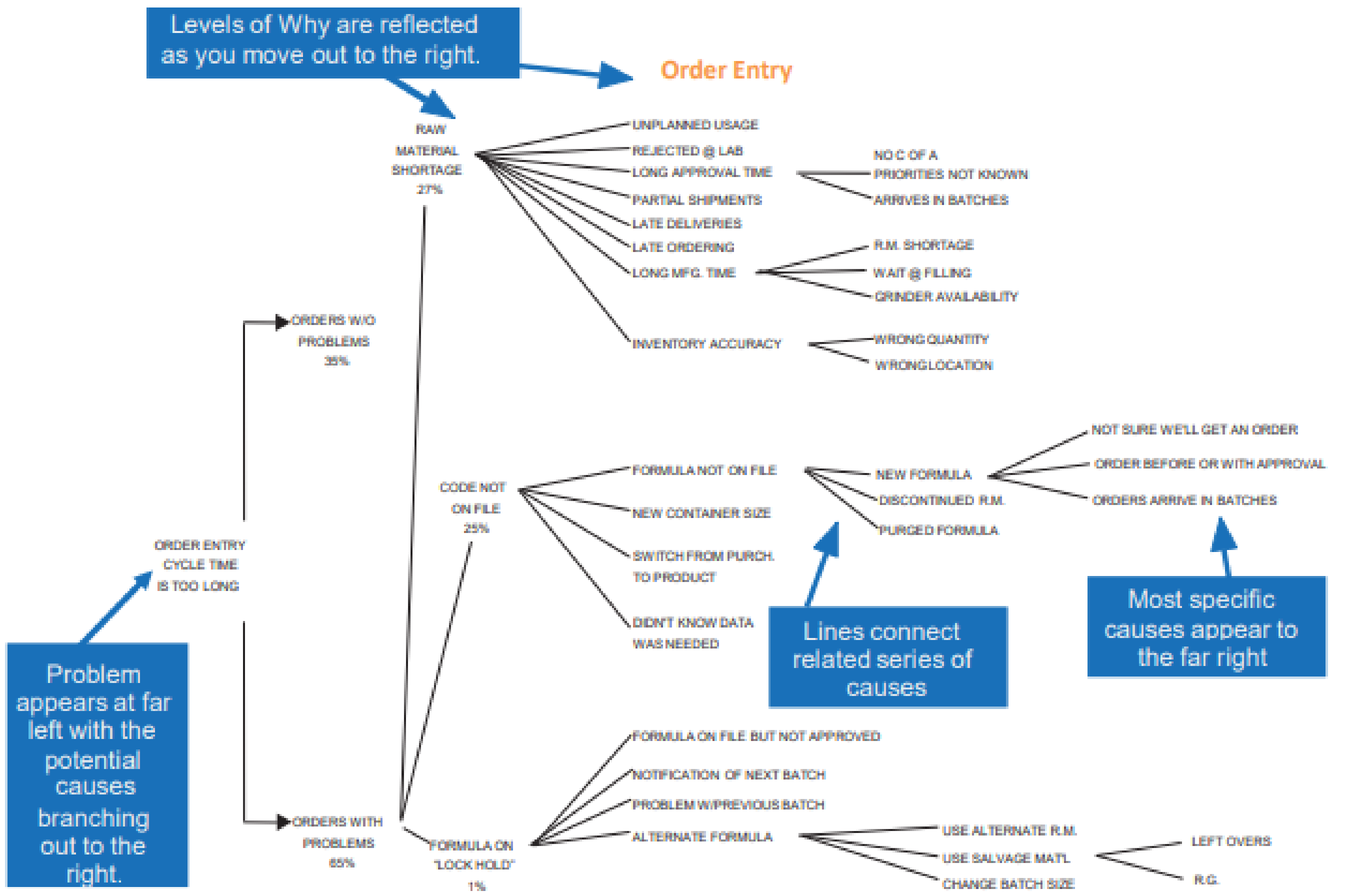


Tree Diagram – How it should be done

Another way to find structure in potential causes is to use a tree diagram, which is a tool used to arrange related ideas in sequence from broad and general to narrow and specific

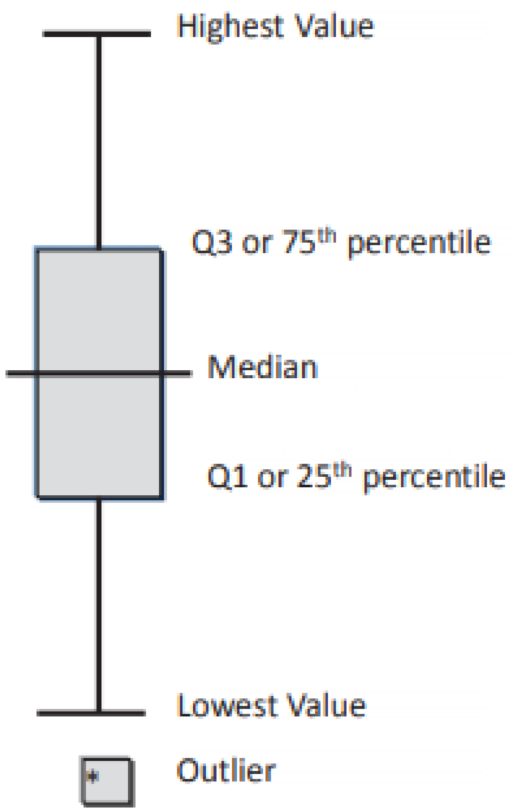


Tree Diagram – How it should be done



Data visualization - Box Plot

A box plot summarizes information about the shape, dispersion and centering; it also helps spot outliers in the data set

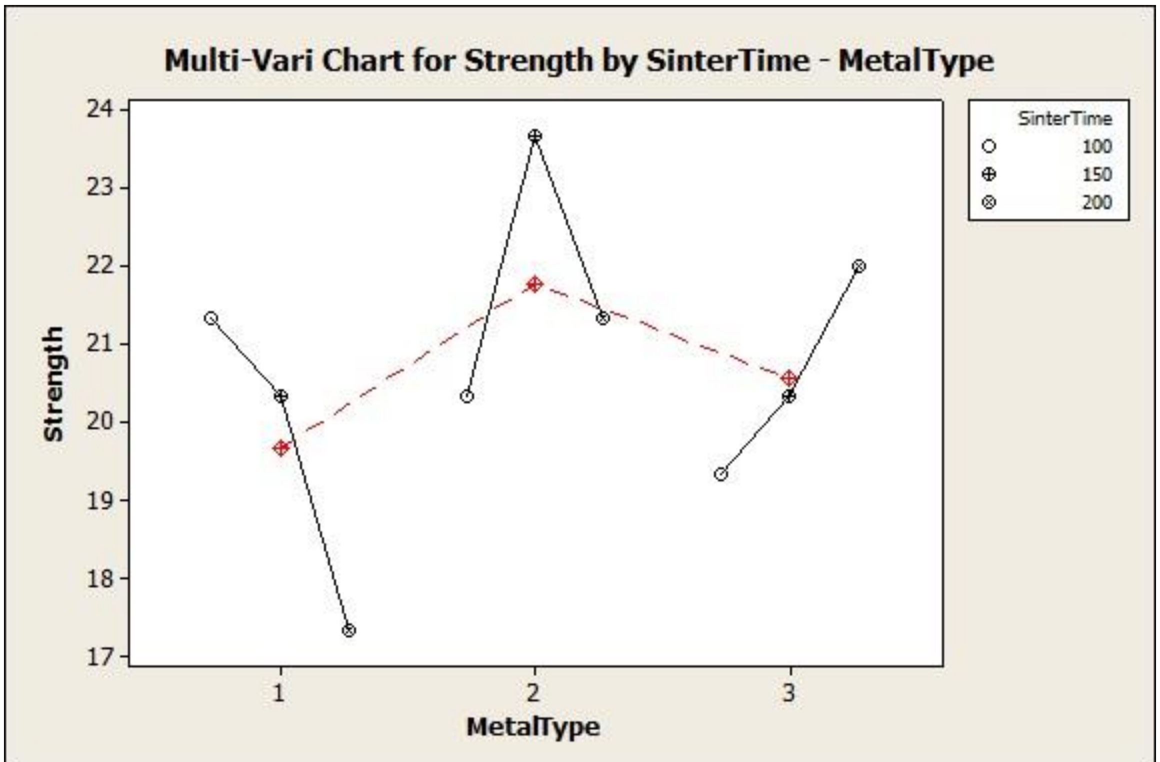


- •BOX – represents the middle 50% values of the data set
- •MEDIAN – represents the point for which 50% of the data points are above and 50% of the data points are below in the line
- •Q1 and Q3 – Q1 represents the point for which 25% of the data points are below and 75% of the data points are above in the line; Q3 represents 75% of the data points are below and 25% of the data points are above in the line
- •AESTRIX – represents an outlier and is a point which does not belong to the family of Xs in the data; beyond Q3+1.5(Q3-Q1) or Q1-1.5(Q3-Q1)
- •LINES – These vertical lines represent a whisker which joins Q1 or Q3 with the farthest data point but other than the outlier ; Maximum Value = Q3+1.5(Q3-Q1) ; Minimum Value = Q1-1.5(Q3-Q1)

 Six Sigma Black Belt Course Accredited by TUV, ICFQ, IASSC

Data visualization – Multi Vari charts

Multi-Vari Chart graphically displays patterns of variation. It is used to identify possible Xs or families of variation, such as variation within a subgroup, between subgroups, or over time



Hypothesis testing

A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data. Hypothesis testing is the process of using statistics to determine the probability that a specific hypothesis is true or false.

To prove that a hypothesis is true, or false, with absolute certainty, we would need absolute knowledge. That is, we would have to examine the entire population.
Instead, hypothesis testing concerns on how to use a random sample to judge if it is evidence that supports or not the hypothesis.

This text book definition may sound a bit complex ….

Lets understand this with a help of a real life example

Hypothesis testing

Scenario 1

Operations Manager of Tier 1 automobile component manufacturing company makes a statement in weekly Production Planning and Control meeting that
“ Morning shift take less time to produce one unit compared to the night shift… hence morning shift is more productive compared to night shift”

What operations Manager trying to prove is called “Real Life Hypothesis” in this case, Morning shift is performing better than the night shift

If we try and convert this “Real Life Hypothesis” in to “Statistical Hypothesis” we will rephrase as “The Average handling time of morning shift is lesser than the average handling time of night shift”

Hypothesis testing

Now we have perfect opportunity to perform a hypothesis testing

There is no difference… every thing is good and no problem is called Null Hypothesis (H o)

“The Average handling time of morning shift is same as the average handling time of night shift”

What we want to prove is called Alternate Hypothesis (H a)
“The Average handling time of morning shift is lesser than the average handling time of night shift”

Test result should either tell us Ho is true or Ha is true

Hypothesis testing

Step1 : Understand the scenario
Step2 : Frame "real life Hypothesis"
Step3 : Convert "real life Hypothesis" in to "statistical Hypothesis"
Step4 : This "statistical Hypothesis" is called "Alternate Hypothesis" (Ha)
Step5 : Based on "Alternate Hypothesis" (Ha) frame "Null Hypothesis" (Ho)
Step6 : Collect sample
Step7 : Perform test statistics
Step8 : Based on the test result either go with Ho or Ha
Step9 : Convert test result in to Business context

Going Back in time

In a courtroom, the person is assumed innocent until proven guilty. In a hypothesis test, we assume the null hypothesis is true until the data proves otherwise

It is only later, in 1933, that "hypothesis tests" were invented by two statisticians: Polish Jerzy Neyman and British Egon Pearson

Hypothesis testing

Now we have decided to test this hypothesis… Operations manager (OM) requested the Team Lead (TL) of both the shift to collect AHT of 60 sample each from their respective teams

After few days both the TLs reverted back to the OM with 60 samples of AHT and the corresponding sample averages are
Morning shift : 17 Min 42 Sec
Night Shift : 20 Min 35 Sec

Most of us would have made your decision by looking at these values but failed to notice

Question to audience is … Will you go with Ho or Ha

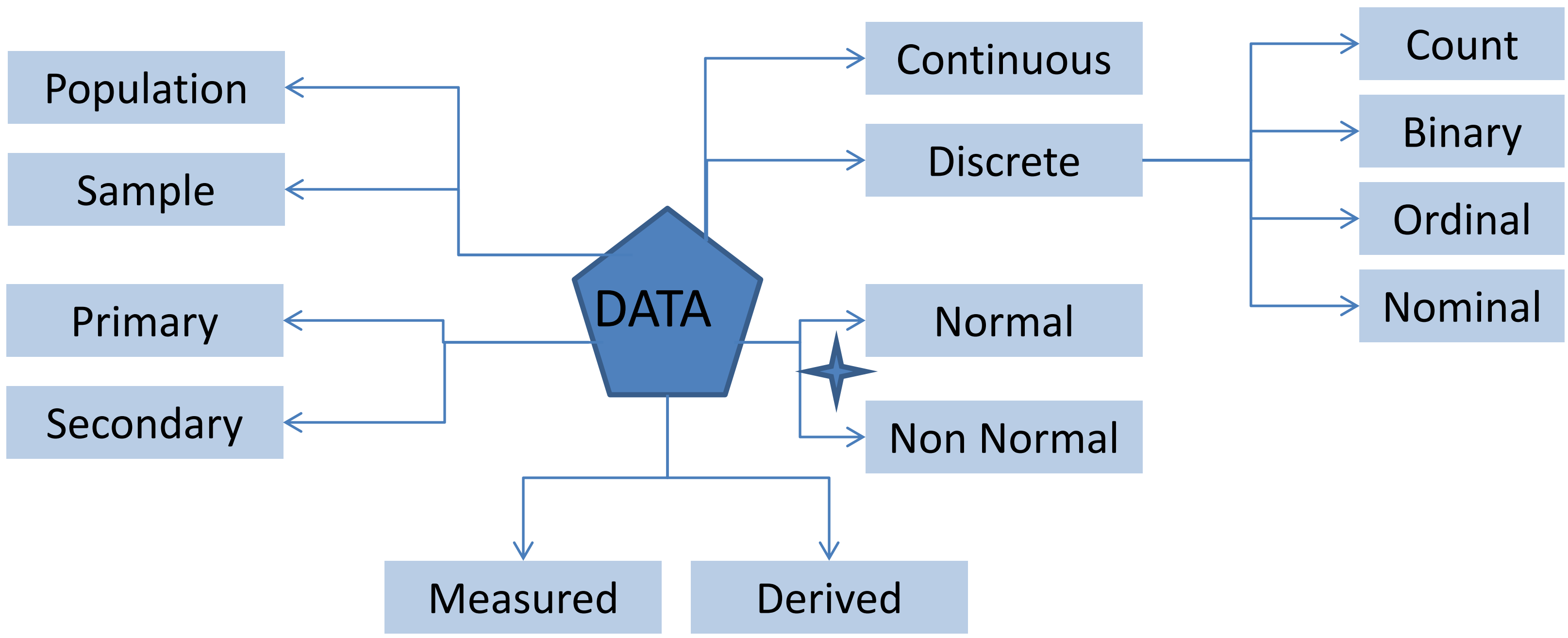
Null Hypothesis (H o) : “The Average handling time of morning shift is same as the average handling time of night shift”

Null Hypothesis (H a) : “The Average handling time of morning shift is lesser than the average handling time of night shift”

Hypothesis testing

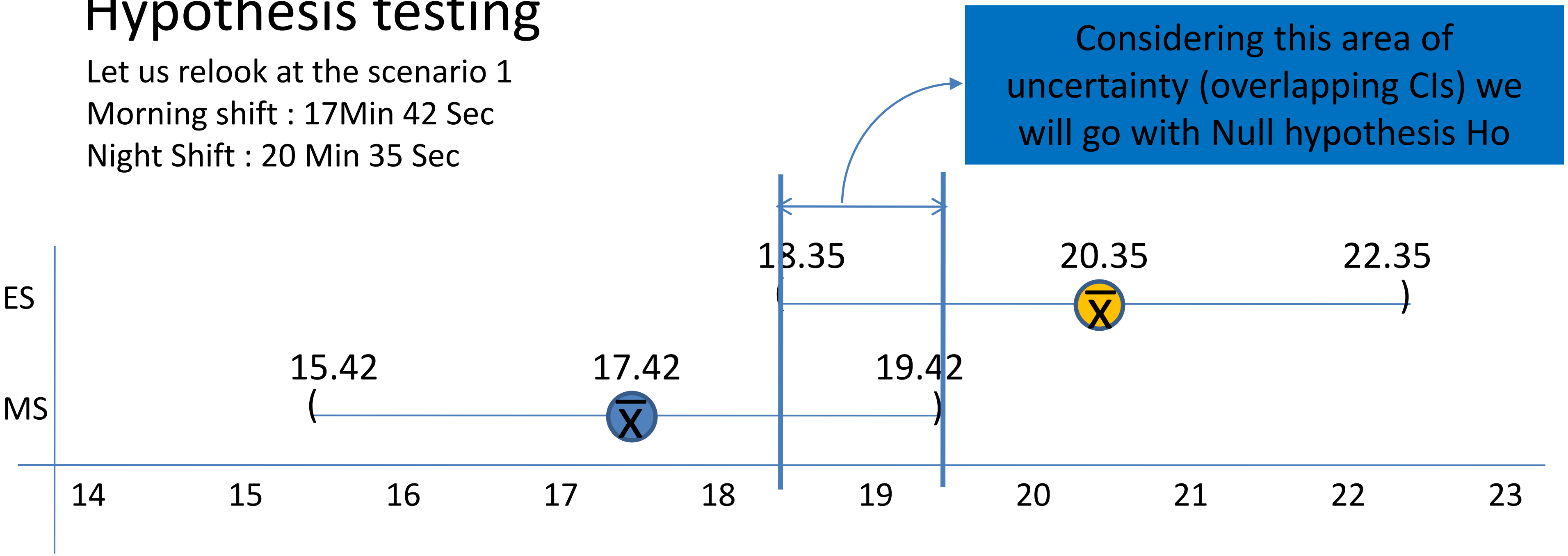
What is the relevance of "Sample average" in previous slide?

Understanding the DATA under study is very important in hypothesis testing



Hypothesis testing

Let us relook at the scenario 1
Morning shift : 17Min 42 Sec
Night Shift : 20 Min 35 Sec



Question to audience is … Will go with Ho or Ha

Null Hypothesis (H o) : "The Average handling time of morning shift is same as the average handling time of night shift"

Alternate Hypothesis (H a) : "The Average handling time of morning shift is lesser than the average handling time of night shift"

Hypothesis testing

Now we understand confidence interval plays a very important role deciding the out come of Hypothesis testing

Standard Error = $\sqrt{\frac{p(1-p)}{n}}$	Standard Error = $\sigma/\sqrt{n}$
Margin of Error = $2\sqrt{\frac{p(1-p)}{n}}$	Margin of Error = $2\sigma/\sqrt{n}$
Confidence interval is $\hat{p} \pm margin\ of\ error$ $\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}$	Confidence interval is $\overline{X} \pm margin\ of\ error$ $\overline{X} \pm 2\sigma/\sqrt{n}$
Both expected successes and failures at least 10	Variable is normally distributed in the population OR sample size is more than 30

Based on these formulas we can make the following statements

If we take less samples Confidence interval will be wider

If we take more samples Confidence interval will be Narrow

So it becomes very clear that we need to identify the accurate sample size and perform hypothesis testing

Hypothesis testing

What can go wrong while doing Hypothesis testing

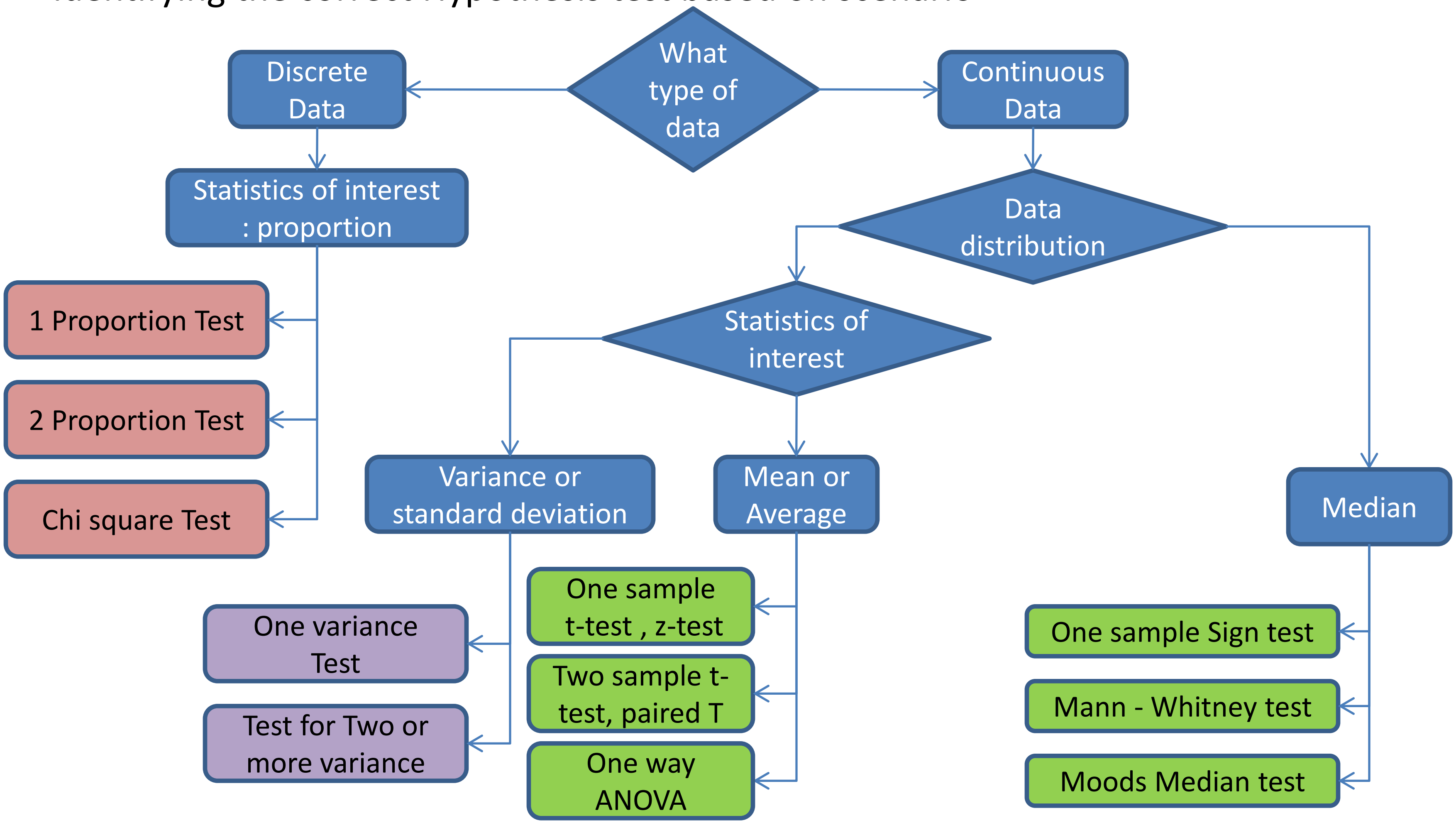
Accept \ Truth	H_o	H_a
H_o	**Correct Decision** $1 - \alpha$ = Confidence	**Type II Error** β
H_a	**Type I Error** α	**Correct Decision** $1 - \beta$ = Power

$1 - \beta$ = Probability of correctly accepting the alternative hypothesis (detecting a change when there is one).

$1 - \alpha$ = Probability of correctly accepting the null hypothesis (detecting no change when there is none).

Hypothesis testing

Identifying the correct Hypothesis test based on scenario



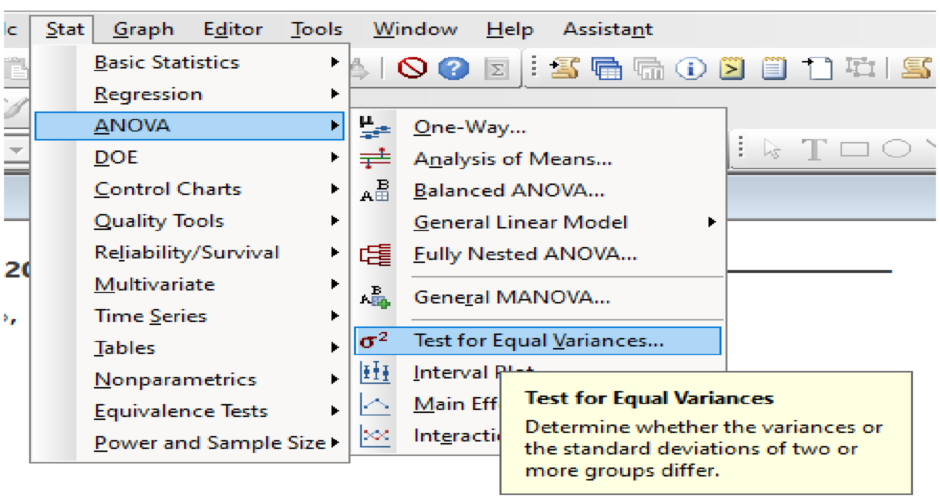
Hypothesis testing - Statistics of interest : Variance

Pre Requisites	
Data Type	Continuous
Distribution	Normal

Test for two or more variance

Determine whether the variance or the standard deviation of two or more groups are different

Business Scenario: A vegetable oil company procure a “Chemical -x” one of the important ingredients from two different suppliers. As Quality manager you are tasked to find if the pH value of the chemical of both the suppliers have same level of variation



Minitab exercise

Refer : Test for Two variance

Hypothesis testing - Statistics of interest : Variance

Pre Requisites	
Data Type	Continuous
Distribution	Normal

Test for two or more variance

Determine whether the variance of two or more groups are different

Null hypothesis: All variances are equal
Alternative hypothesis: At least one variance is different

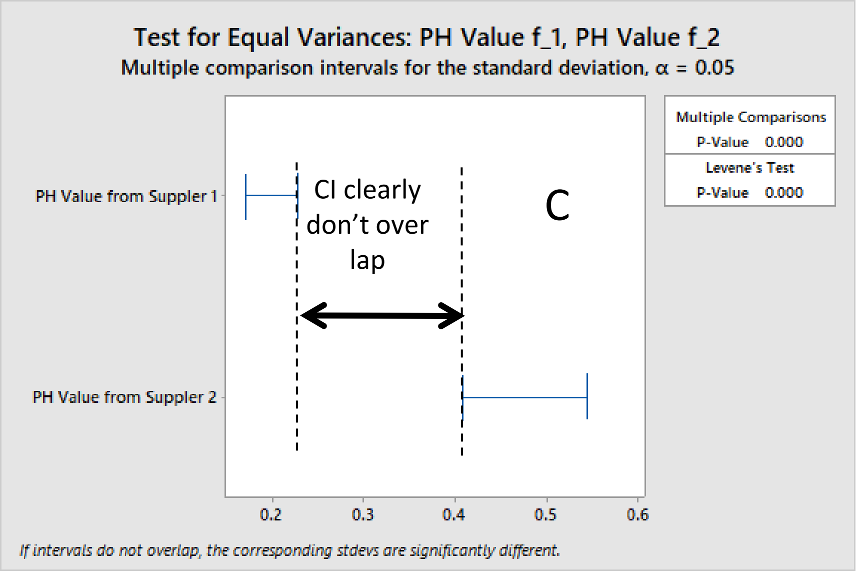
95% Bonferroni Confidence Intervals for Standard Deviations

Sample	N	StDev	CI
PH Value from Suppler 1	100	0.196350	(0.169943, 0.232061)
PH Value from Suppler 2	100	0.467028	(0.392291, 0.568752)

Individual confidence level = 97.5%

Tests

Method	Test Statistic	P-Value
Multiple comparisons	34.49	0.000
Levene	37.08	0.000



P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

Hypothesis testing - Statistics of interest : Variance

Pre Requisites	
Data Type	Continuous
Distribution	Normal

Test for two or more variance

Determine whether the variance of two or more groups are different

Null hypothesis: All variances are equal
Alternative hypothesis: At least one variance is different

95% Bonferroni Confidence Intervals for Standard Deviations

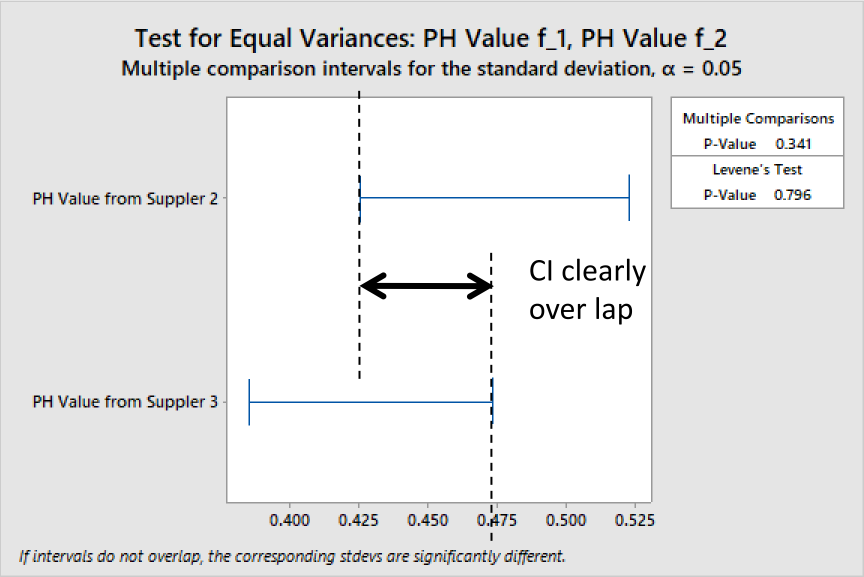
Sample	N	StDev	CI
PH Value from Suppler 2	100	0.467028	(0.392291, 0.568752)
PH Value from Suppler 3	100	0.422790	(0.372456, 0.490930)

Individual confidence level = 97.5%

Tests

Method	Test Statistic	P-Value
Multiple comparisons	0.91	0.341
Levene	0.07	0.796

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis



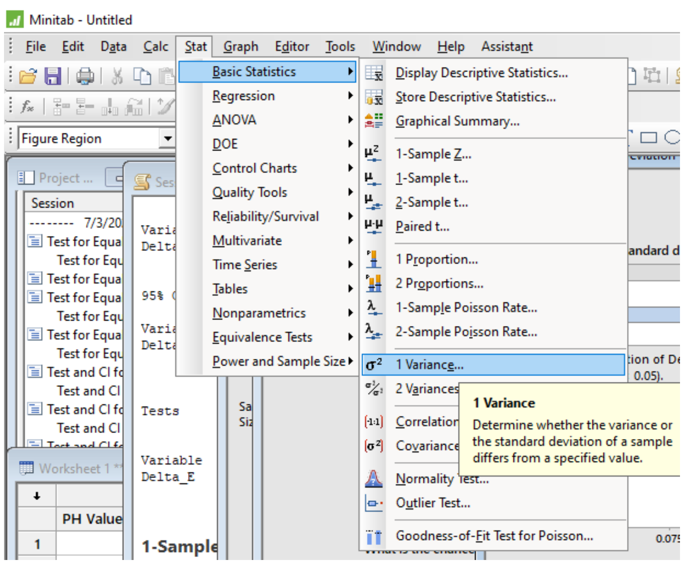
Hypothesis testing - Statistics of interest : Variance

Pre Requisites	
Data Type	Continuous
Distribution	Normal

One variance test

Determine whether the variance or the standard deviation of a sample is different from a specified value

Business Scenario: An Automobile manufacturing company procure paint from leading MNC paint company. As per the contract the acceptable level of Delta – E variation is 0.04 with a mean of 1.2, As procurement manager you are tasked to check if the variation is with in the 0.04 level as per contract



Minitab exercise
Refer : One variance test

Hypothesis testing - Statistics of interest : Variance

Pre Requisites	
Data Type	Continuous
Distribution	Normal

One variance test

Determine whether the variance or the standard deviation of a sample is different from a specified value

Null hypothesis σ = 0.04
Alternative hypothesis σ > 0.04

The chi-square method is only for the normal distribution. The Bonett method is for any continuous distribution.

Statistics

Variable	N	StDev	Variance
Delta_E	60	0.0923	0.00851

95% Confidence Intervals

Variable	Method	CI for StDev	CI for Variance
Delta_E	Chi-Square	(0.0782, 0.1125)	(0.00612, 0.01266)
	Bonett	(0.0803, 0.1096)	(0.00645, 0.01200)

Tests

Variable	Method	Test Statistic	DF	P-Value
Delta_E	Chi-Square	200.92	59	0.000
	Bonett	—	—	0.000

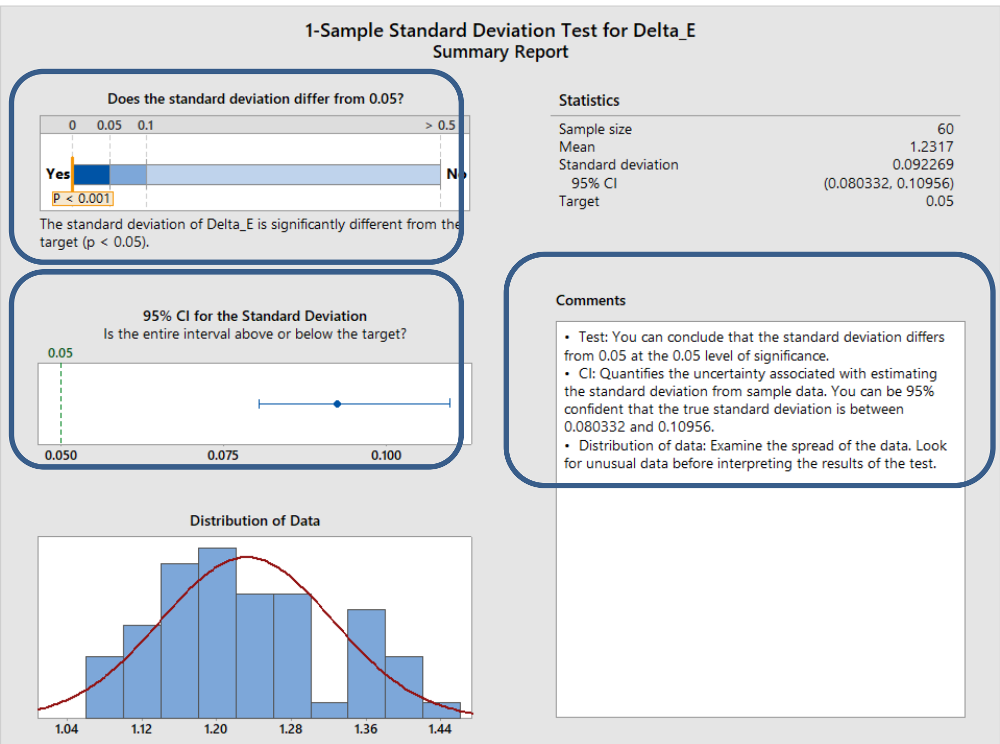
P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

Pre Requisites	
Data Type	Continuous
Distribution	Normal

Hypothesis testing - Statistics of interest : Variance

One variance test

Determine whether the variance or the standard deviation of a sample is different from a specified value



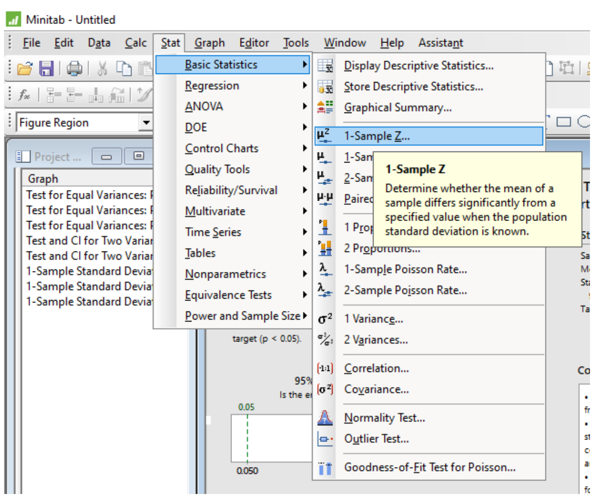
Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

1 Sample Z-test

Determine whether the mean of a sample differs significantly from a specified value when the population standard deviation is known

Business Scenario: As training manager you are asked check if the average call handling time (AHT) of two training teams team a and team b is not greater than 12 min (Current AHT of the operations tem). from operations team we also know the standard deviation of the populations is 1min



Minitab exercise

Refer : 1 Sample Z-test

Hypothesis testing - MEAN

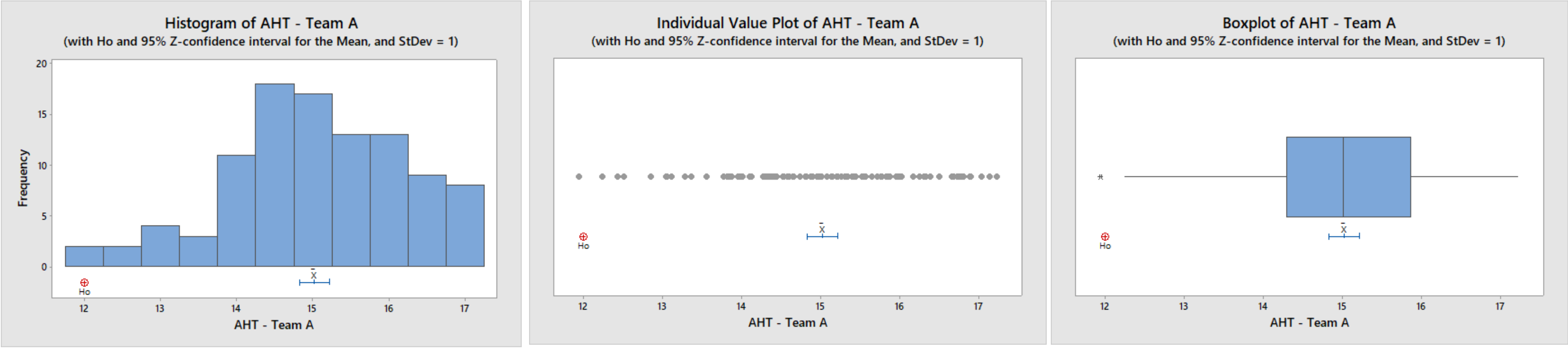
Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

1 Sample Z-test

Null hypothesis µ of Team A = 12 Min

Alternative hypothesis µ of Team A > 12 Min

Variable	N	Mean	StDev	SE Mean	95% CI	Z	P
AHT - Team A	100	15.021	1.170	0.100	(14.825, 15.217)	30.21	0.000



P-Value < 0.05 go with Alternative hypothesis

P-Value > 0.05 go with Null hypothesis

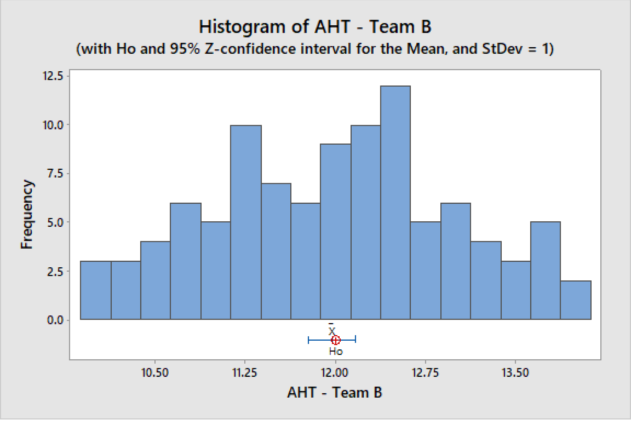
Hypothesis testing - MEAN

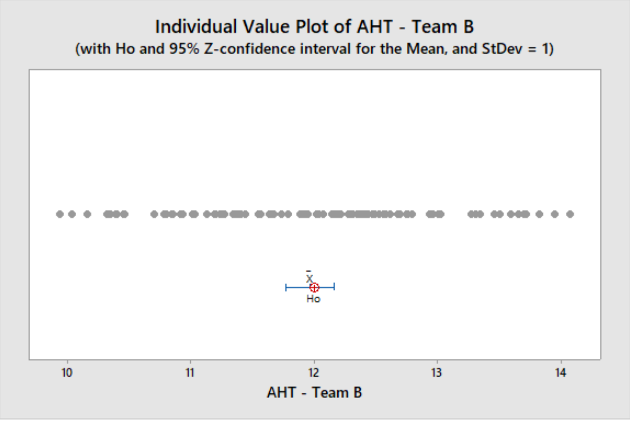
Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

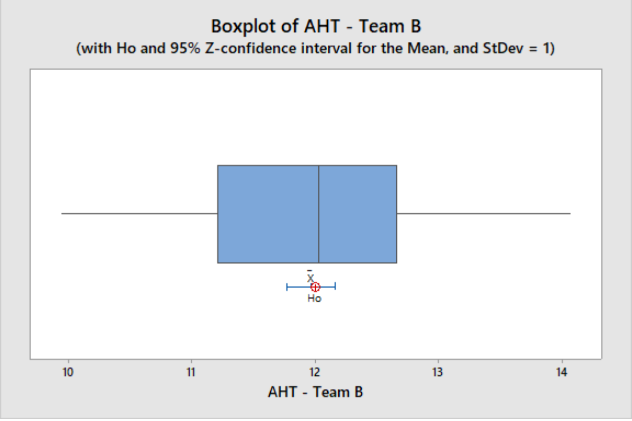
1 Sample Z-test

Null hypothesis μ of Team B = 12 Min
Alternative hypothesis μ of Team B > 12 Min

Variable	N	Mean	StDev	SE Mean	95% CI	Z	P
AHT - Team B	100	11.970	1.017	0.100	(11.774, 12.166)	-0.30	0.767







P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

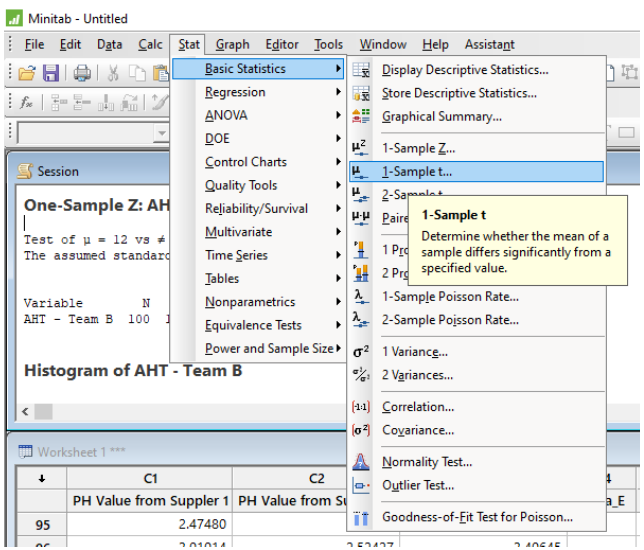
Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

1 Sample t-test

Determine whether the mean of a sample differs significantly from a specified value.

Business Scenario: A Semiconductor making company decided to evaluate a new supplier for Mercury . As a R&D team member you are tasked to check the average viscosity of the liquid metal supplied in different samples meets the requirement of 1.526 cP (centipoise)



Minitab exercise

Refer : 1 Sample Z-test

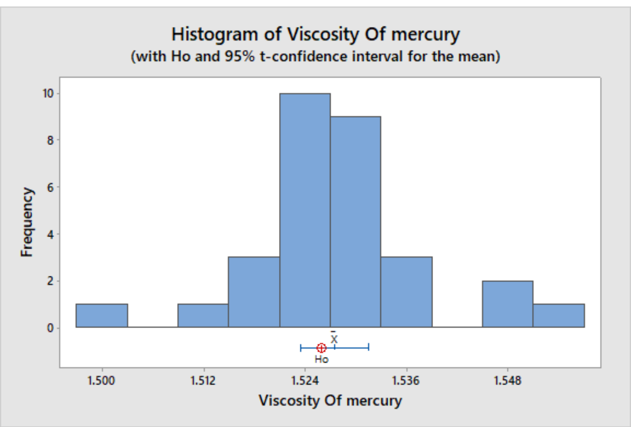
Hypothesis testing - MEAN

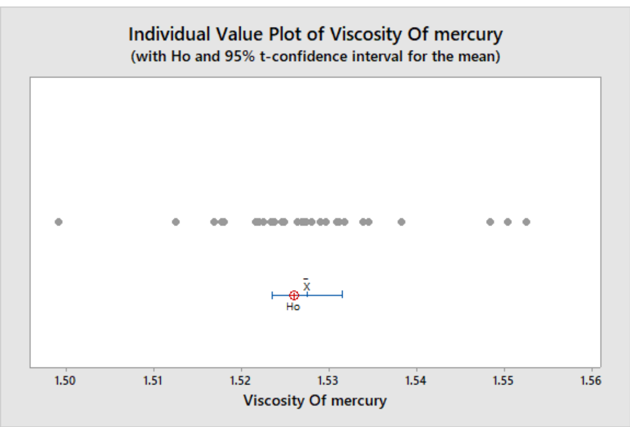
Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

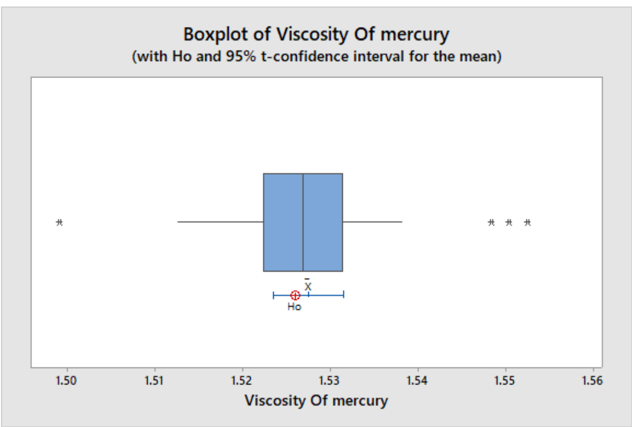
1 Sample t-test

Null hypothesis µ velocity of Mercury supplied by new supplier = 1.526 cP
Alternative hypothesis µ velocity of Mercury supplied by new supplier ≠ 1.526 cP

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Viscosity Of mercury	30	1.52751	0.01072	0.00196	1.52351, 1.53151)	0.77	0.446







P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

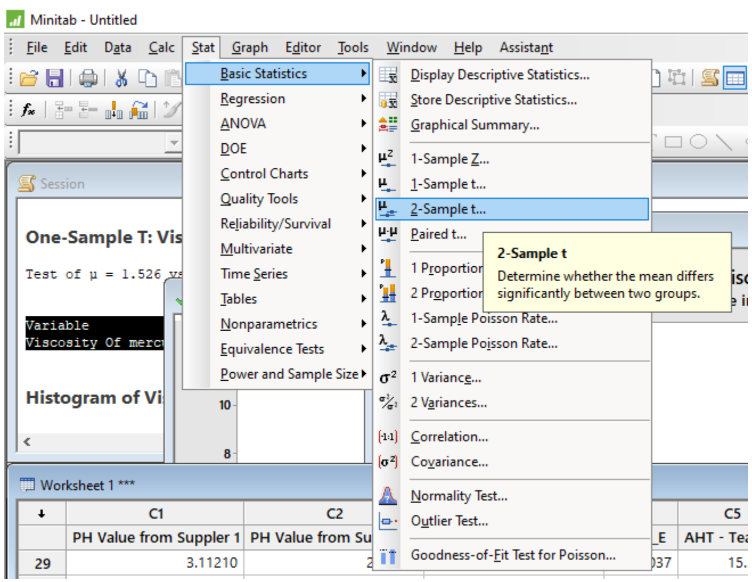
Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

2 Sample t-test

Determine whether the mean differs significantly between two groups

Business Scenario: Operations manager of a contact centre need to identify between team managed by Team leader Mike and Team managed by Team leader David which team have relatively less AHT



Minitab exercise

Refer : 2 Sample T-test

Hypothesis testing - MEAN

2 Sample t-test

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

Two-Sample T-Test and CI: Team - Mike, Team - David
Two-sample T for Team - Mike vs Team – David

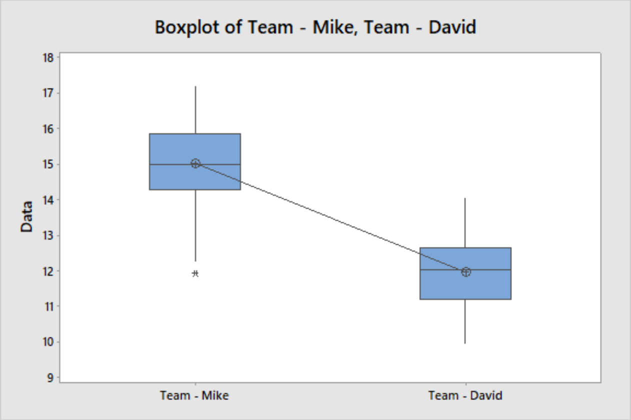
Null hypothesis μ AHT of Team - Mike = μ AHT of Team - David
Alternative hypothesis μ AHT of Team - Mike ≠ μ AHT of Team - David

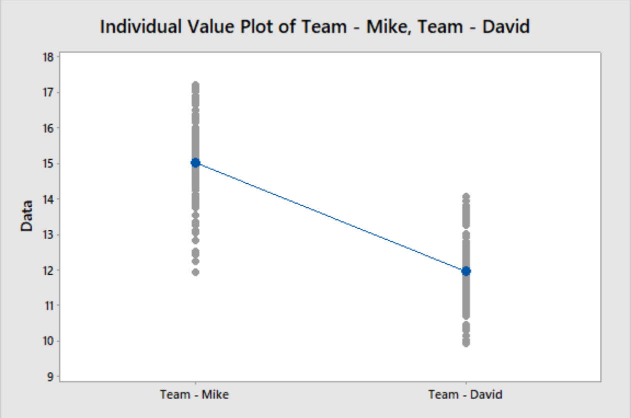
	N	Mean	StDev	SE Mean
Team - Mike	100	15.02	1.17	0.12
Team - David	100	11.97	1.02	0.10

Difference = μ (Team - Mike) - μ (Team - David)
Estimate for difference: 3.051
95% CI for difference: (2.745, 3.357)
T-Test of difference = 0 (vs ≠):

T-Value = 19.68 P-Value = 0.000 DF = 194

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

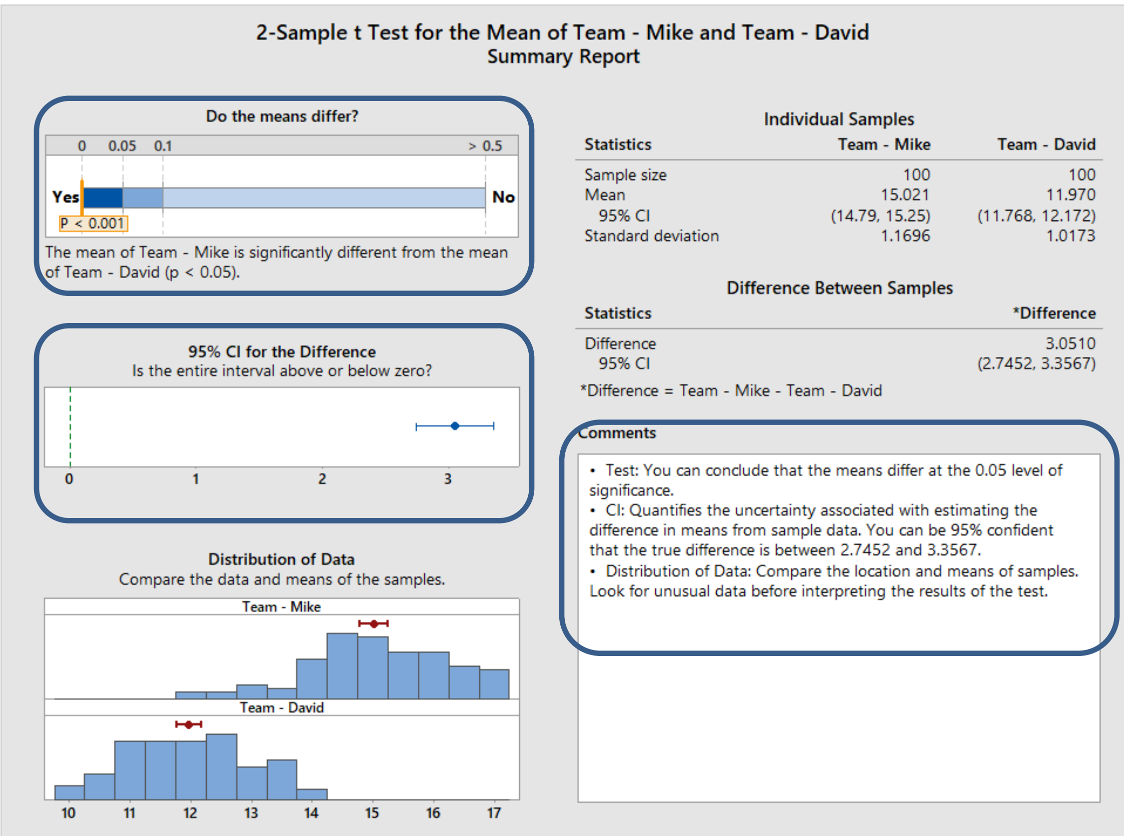




Hypothesis testing - MEAN

2 Sample t-test

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	



2-Sample t Test for the Mean of Team - Mike and Team - David
Summary Report

Do the means differ?

The mean of Team - Mike is significantly different from the mean of Team - David (p < 0.05).

95% CI for the Difference
Is the entire interval above or below zero?

Distribution of Data
Compare the data and means of the samples.

Individual Samples

Statistics	Team - Mike	Team - David
Sample size	100	100
Mean	15.021	11.970
95% CI	(14.79, 15.25)	(11.768, 12.172)
Standard deviation	1.1696	1.0173

Difference Between Samples

Statistics	*Difference
Difference	3.0510
95% CI	(2.7452, 3.3567)

*Difference = Team - Mike - Team - David

Comments

- Test: You can conclude that the means differ at the 0.05 level of significance.
- CI: Quantifies the uncertainty associated with estimating the difference in means from sample data. You can be 95% confident that the true difference is between 2.7452 and 3.3567.
- Distribution of Data: Compare the location and means of samples. Look for unusual data before interpreting the results of the test.

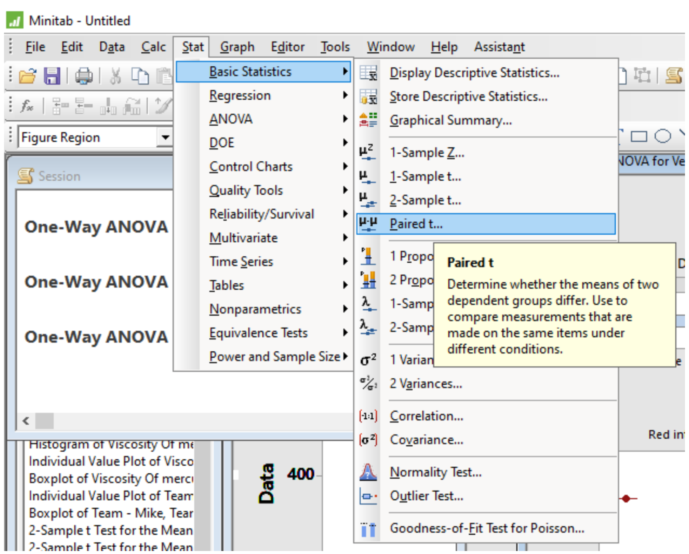
Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

Hypothesis testing - MEAN

Paired t-test

Used to compare measurements that are made on the same item under different conditions

Business Scenario: An F1 teams R&D department want to check if response time of moving from 0-100 km differ based on Air or Nitrogen in the tyres



Minitab exercise

Refer : Paired t-test

Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

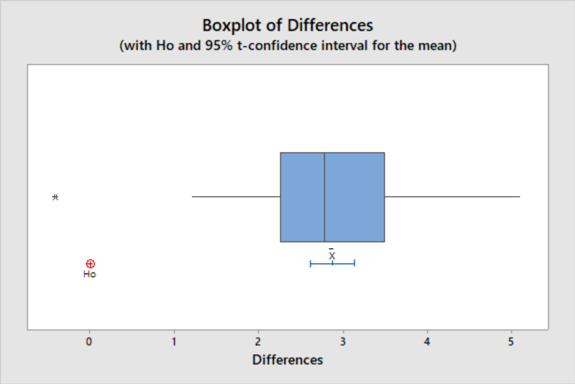
Paired t-test

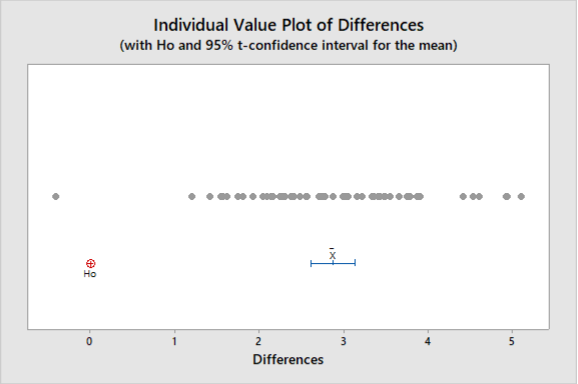
Paired T for Air - Nitrogen

Null hypothesis μ Response time of Air = μ Response time of Nitrogen
Alternative hypothesis μ Response time of Air ≠ μ Response time of Nitrogen

	N	Mean	StDev	SE Mean
Air	60	11.9693	0.7450	0.0962
Nitrogen	60	9.0957	0.7070	0.0913
Difference	60	2.874	1.011	0.130

95% CI for mean difference: (2.613, 3.135)
T-Test of mean difference = 0 (vs ≠ 0):
T-Value = 22.02
P-Value = 0.000





P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

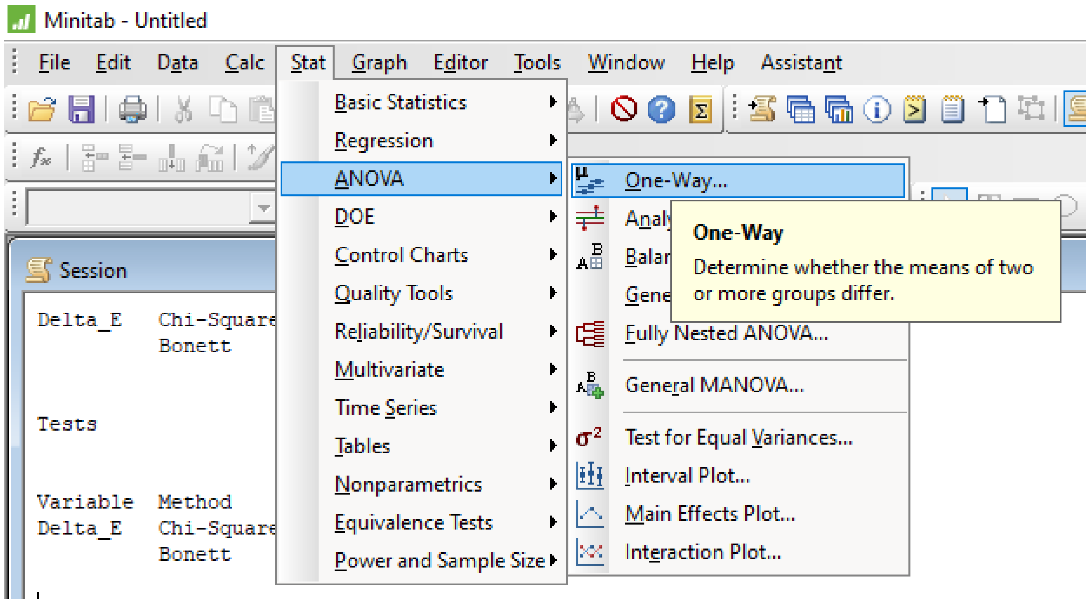
Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

ANOVA

Determine whether the means of two or more groups differs significantly

Business Scenario: Machine shop manager want to check the tensile strength of 10mm steal rods procured form three different sources, to see if all the roads are of same strength or is there any difference (Acceptable level is 400 to 415 PSI)



Minitab exercise

Refer : ANOVA

Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

ANOVA

Null hypothesis All means are equal
Alternative hypothesis At least one mean is different
Significance level $\alpha = 0.05$

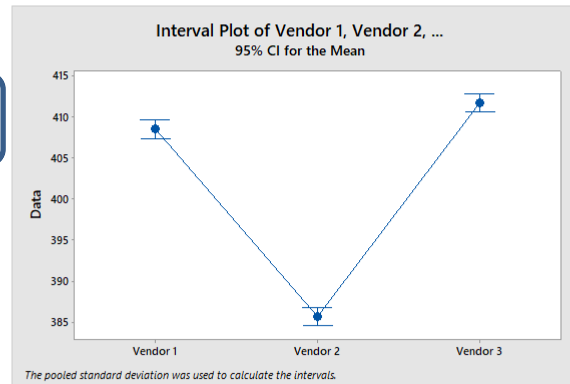
Means

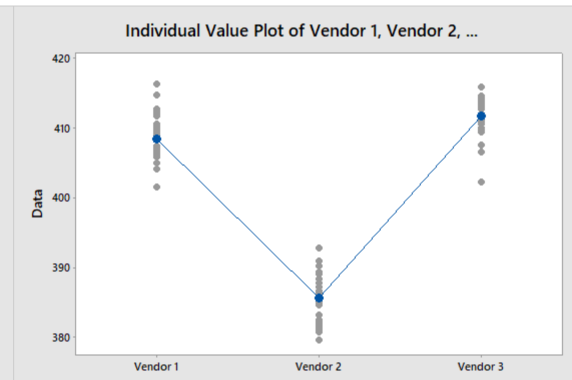
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	2	12051.5	6025.74	643.69	0.000

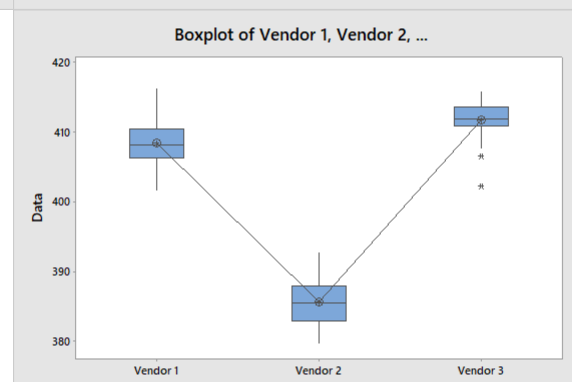
Factor	N	Mean	StDev	95% CI
Vendor 1	30	408.463	3.231	(407.352, 409.573)
Vendor 2	30	385.667	3.230	(384.556, 386.777)
Vendor 3	30	411.654	2.686	(410.543, 412.764)

Pooled StDev = 3.05961

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis



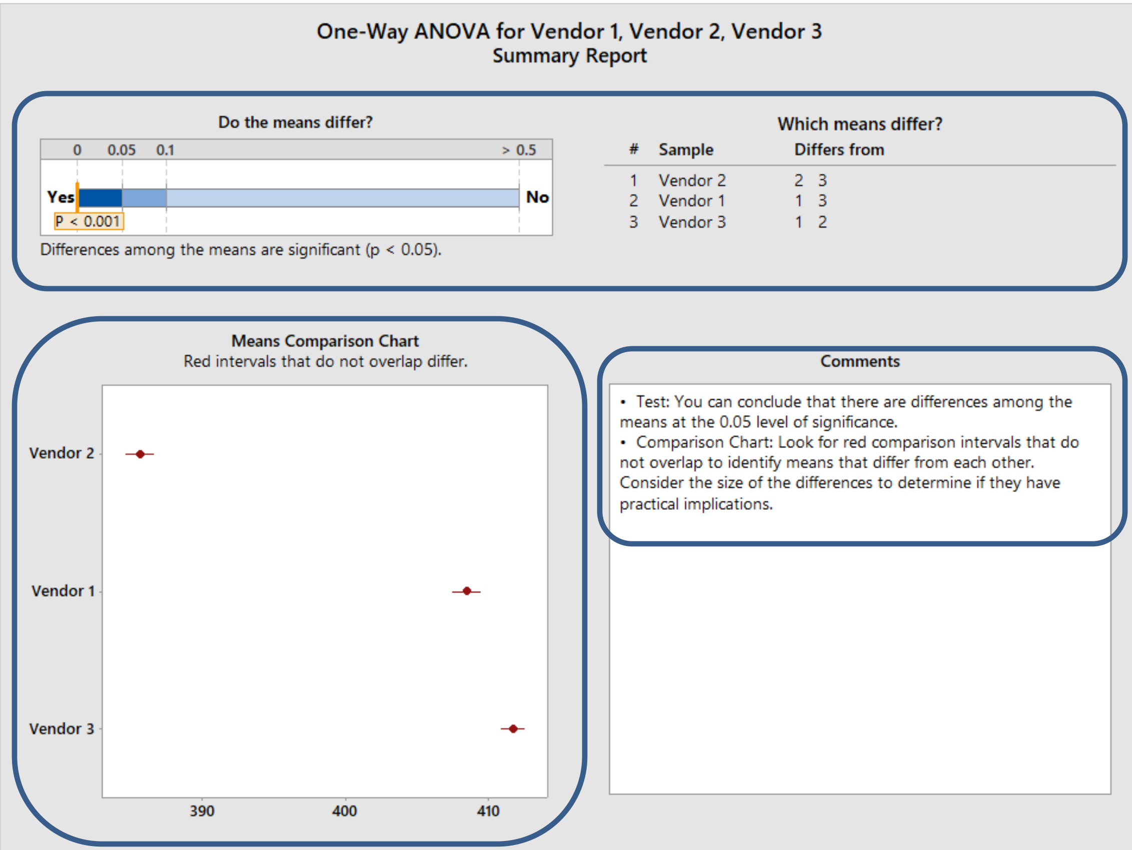




Hypothesis testing - MEAN

Pre Requisites	
Data Type	Continuous
Distribution	Normal
Equal Variance (More than 1 sample data set)	

ANOVA



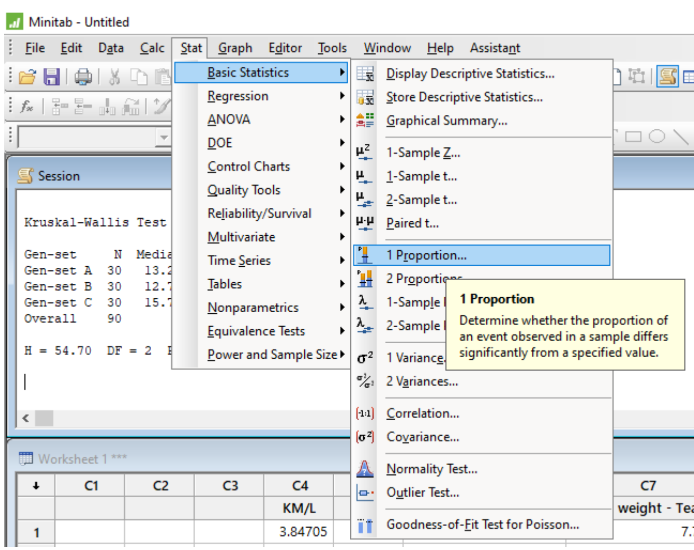
Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

1 Proportion test

Determine whether the Proportion of a event observed in a sample differs from a specified value

Business Scenario: As a Sales Manager you need to statistically prove to potential customer that your machine will produce less defects compared to current defect rate of 6% in the customer shop floor



Minitab exercise

Refer : 1 proportion test

Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

1 Proportion test

Null hypothesis : Proportion defective of new machine A = 6%
Alternative hypothesis : Proportion defective of new machine A < 6%

Test of p = 0.06 vs p ≠ 0.06

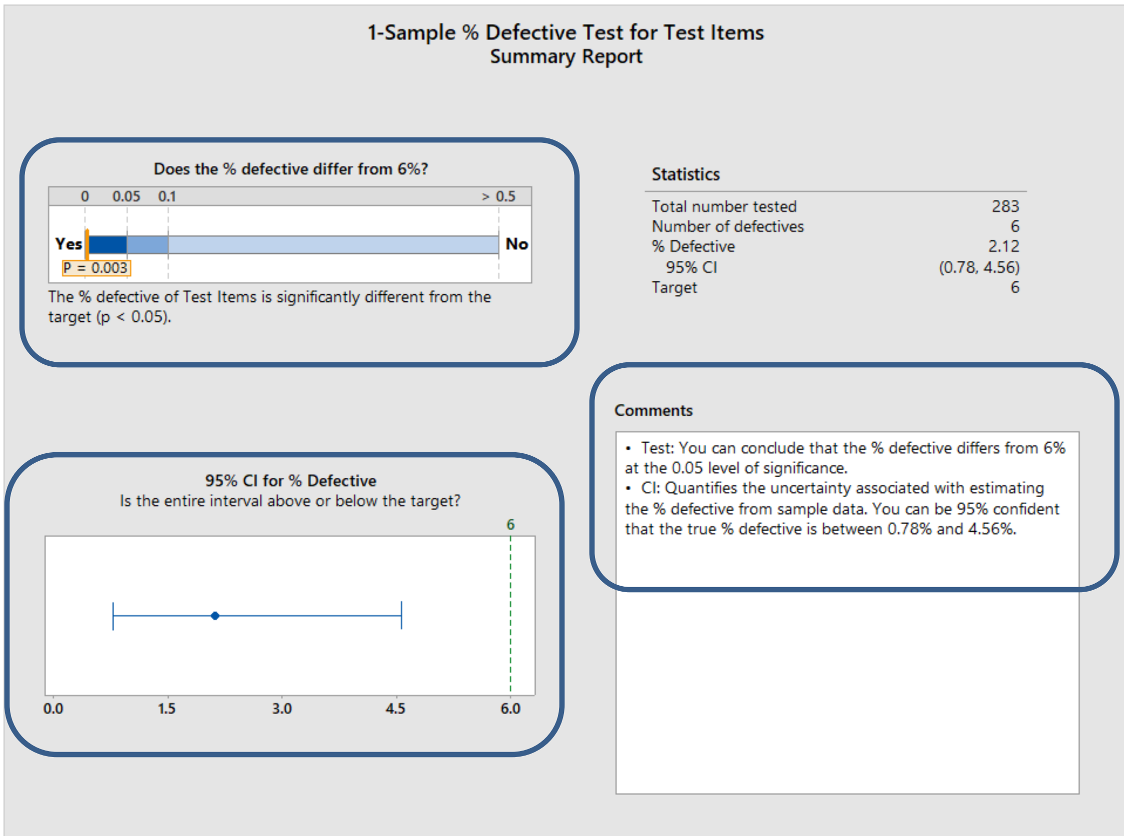
Sample	X	N	Sample p	Exact 95% CI	P-Value
1	6	283	0.021201	(0.007819, 0.045573)	0.003

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

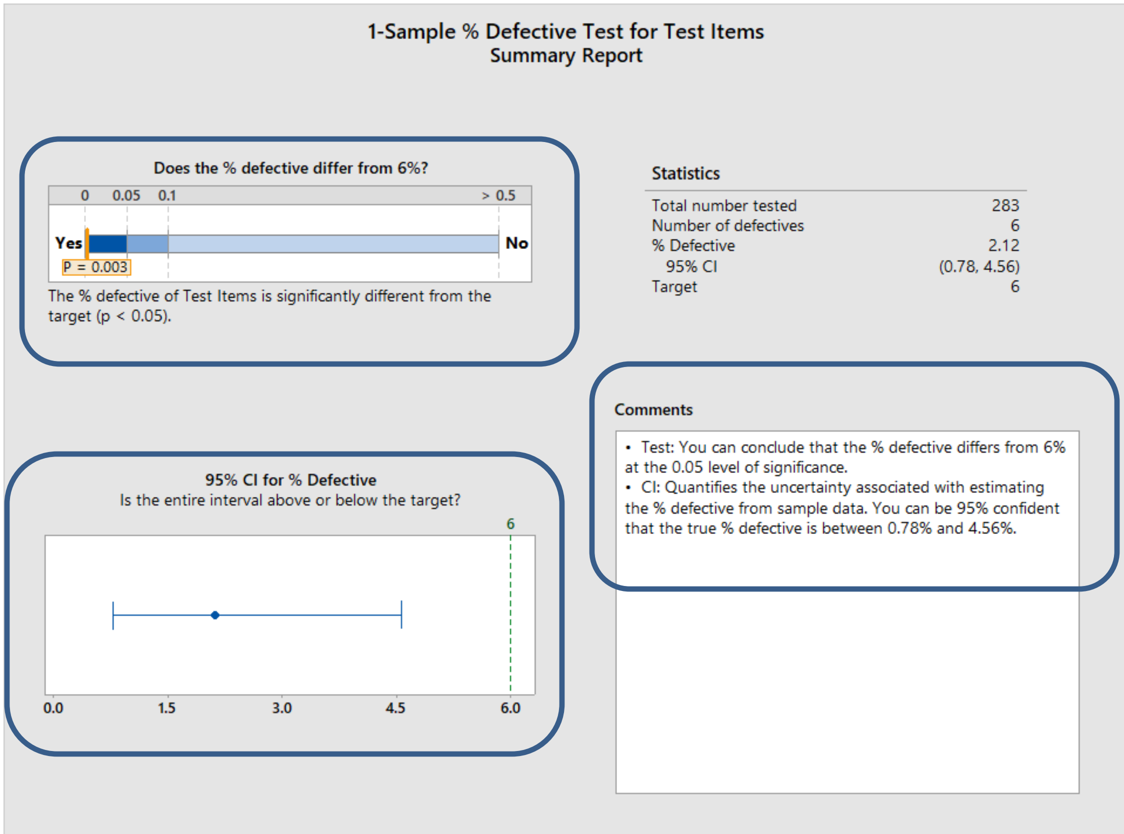
1 Proportion test



Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

1 Proportion test



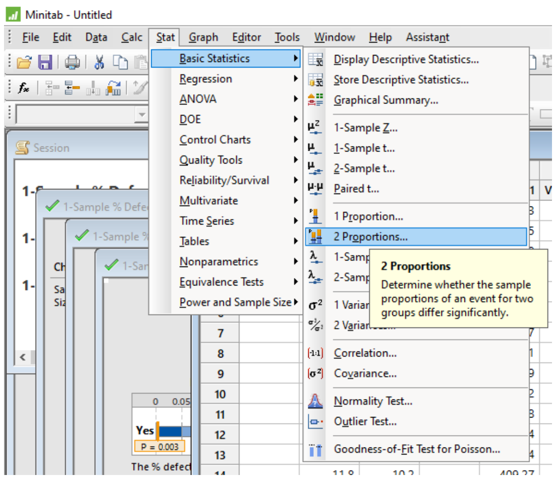
Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

Hypothesis testing - Proportion

2 Proportion test

Determine whether the sample Proportion of a event for two groups differs significantly

Business Scenario: As a procurement manager you have to chose between two suppliers ("A" and "B") of spark plugs, comparing their defective %. Requirement is spark plug should ignite the engine in the first attempt.



Minitab exercise

Refer : 2 proportion test

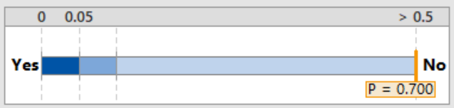
Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

2 Proportion test

2-Sample % Defective Test for Group 1 vs Group 2
Summary Report

Do the % defectives differ?



The % defective of Group 1 is not significantly different from the % defective of Group 2 (p > 0.05).

Individual Samples

Statistics	Group 1	Group 2
Total number tested	153	283
Number of defectives	3	4
% Defective	1.96	1.41
95% CI	(0.41, 5.62)	(0.39, 3.58)

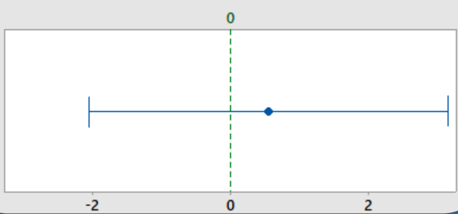
Difference Between Samples

Statistics	*Difference
Difference	0.55
95% CI	(-2.04, 3.14)

*Difference = Group 1 - Group 2

95% CI for the Difference

Is the entire interval above or below zero?



Comments

- Test: There is not enough evidence to conclude that the % defectives differ at the 0.05 level of significance.
- CI: Quantifies the uncertainty associated with estimating the difference from sample data. You can be 95% confident that the true difference is between -2.04% and 3.14%.

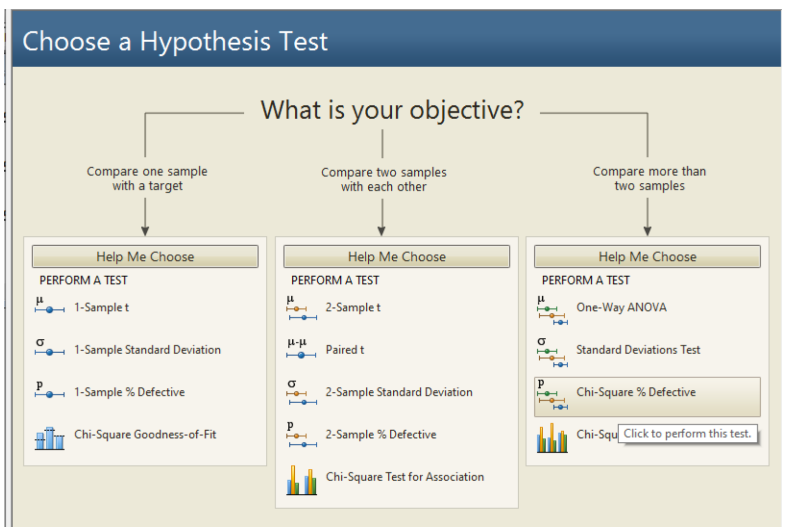
Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

Chi - Square Proportion test

Determine whether the sample Proportion of a event for more two groups differs significantly

Business Scenario: As a procurement manager you have to chose between two suppliers ("A" , "B" and "c") of spark plug comparing their defective %. Requirement is spark plug should ignite the engine in the first attempt.



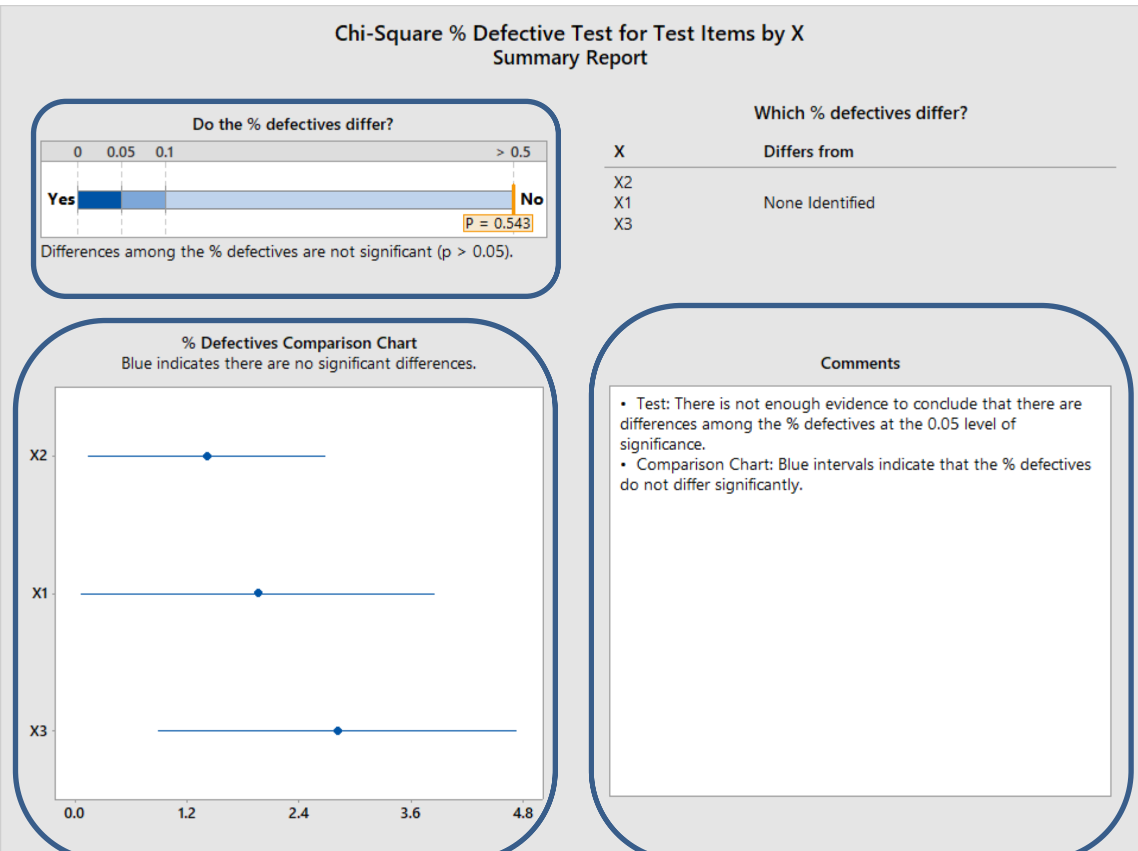
Minitab exercise

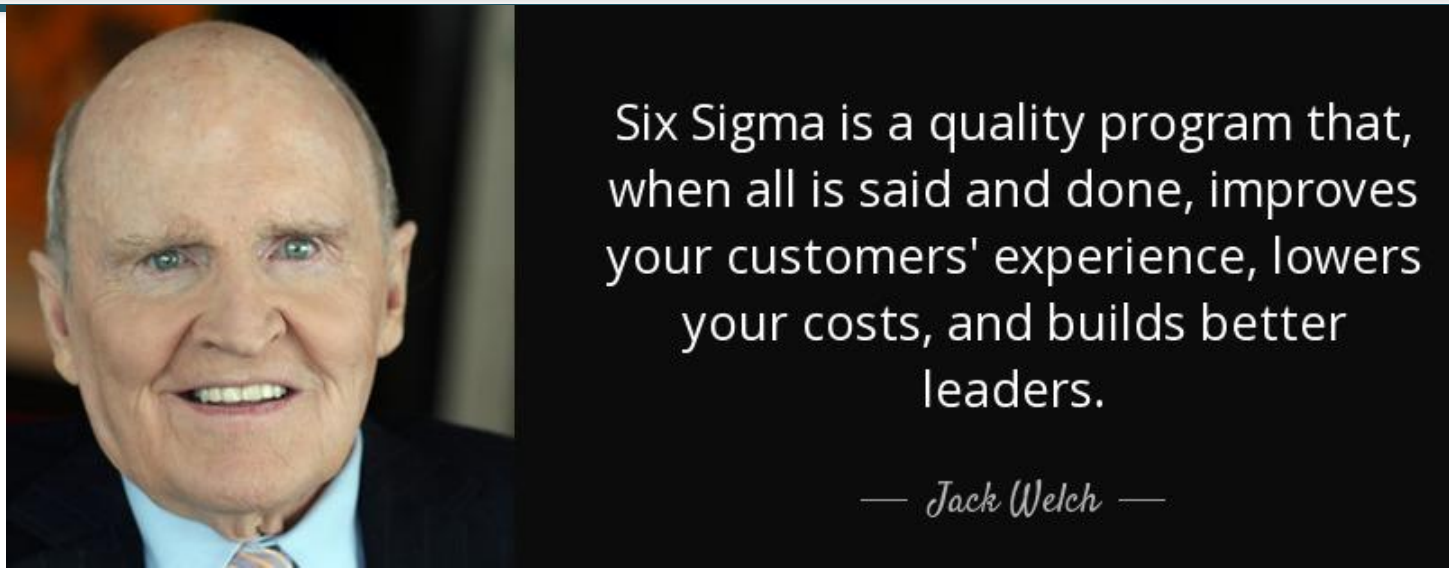
Refer : 1 proportion test

Hypothesis testing - Proportion

Pre Requisites	
Data Type	Discrete
Distribution	Binomial (defective)

Chi - Square Proportion test




Six Sigma is a quality program that, when all is said and done, improves your customers' experience, lowers your costs, and builds better leaders.
— Jack Welch —