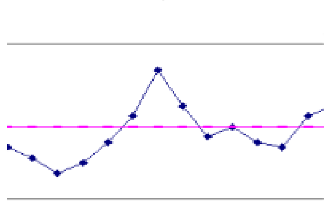
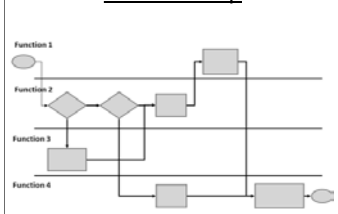
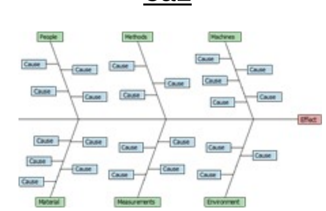
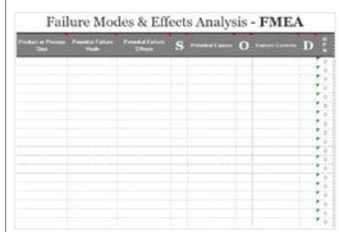
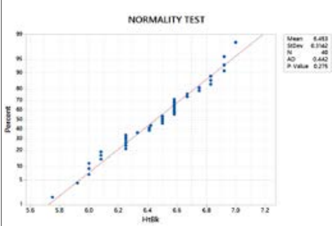
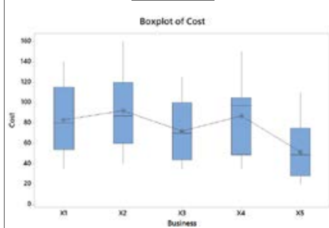
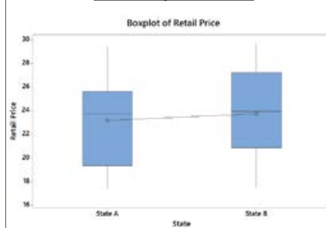
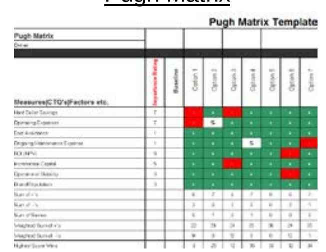
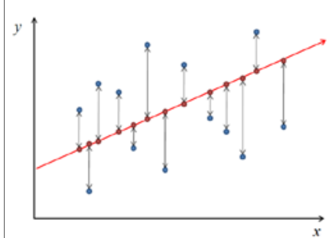
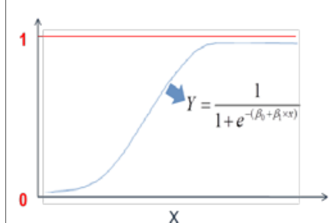
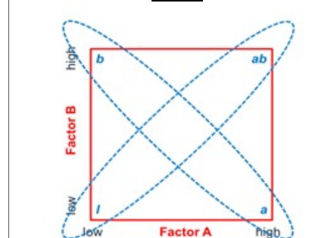
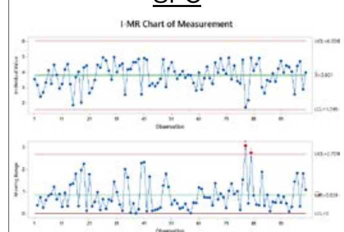
LEAN SIX SIGMA

Quick Reference Guide

™

Lean Sigma Corporation™

Lean Six Sigma DMAIC Roadmap

	Purpose	Key Tools				Key Outputs
Define	To establish a quantified problem statement, objective and business case that will become the foundation to your Six Sigma project. Conduct stakeholder analysis, select team members and kick-off your project.	Primary Metric 	Process Map 	Project Charter 	Project Plan 	* Process Map * Gather VOC * Translate VOC to CTQ's * QFD/HOQ * COPQ * Primary & Secondary Metrics * Establish Project Charter * Stakeholder Analysis * Team Selection * Project Plan
Measure	Refine your understanding of the process. Assess process capability relative to customer specifications. Validate measurement systems. Brainstorm potential x's.	C&E 	SIPOC 	FMEA 	Cpk 	* Early Y=f(x) Hypothesis * Detailed Process Map * SIPOC * Cause & Effect Diagram * Cause & Effect Matrix * FMEA * Basic Statistics * Normality Test * Capability Analysis * Gage R&R
Analyze	Conduct data collection and planned studies in order to eliminate non-critical x's and validate critical x's. Establish a stronger and quantified Y=f(x) equation.	Normality Test 	ANOVA 	2 Sample t-test 	Equal Variances 	* Narrowed Y=f(x) * 1 & 2 Sample t-tests * 1 & 2 Proportions tests * Equal variance tests * Normality tests * ANOVA * Moods Median * Mann Whitney * Paired t-test * Chi-Squared test
Improve	Design, test and implement your new process or product under live operating conditions. Pilot solutions if feasible before broadly deploying expensive improvements or products.	Pugh Matrix 	Linear Regression 	Binary Logistic Regression 	DOE 	* Refined Y=f(x) * Pugh Matrix * Correlation * Simple Linear Regression * Multiple Linear Regression * Binary Logistic Regression * Full Factorial DOE * Fractional Factorial DOE
Control	Plan, communicate, train and implement your product or process solutions. Ensure control mechanisms are established. Use Poke Yoke, visual controls, SOP's and SPC wherever possible.	Control Plan 	SOP's 	Communication Plan 	SPC 	* Control Plan * Training Plan * Refined FMEA * Communication Plan * Standard Operating Procedures * Five-S Audit * Poke Yoke * Visual Controls * Statistical Process Control

Defects per Unit (DPU) = Defects (D) / Units (U): DPU = D/U

Defects per Million Opportunities (DPMO) = [D/(U*O)]*1,000,000
Where 'O' represents opportunities for defects

Rolled Throughput Yield: Multiply the yield of each process step: $Y_1 * Y_2 * Y_3$…

Yield Equation: $Y = e^{-dpu}$

e = 2.718

Yield Estimation Method: 1 – DPU

Program Evaluation & Review Technique (PERT) = [(O + (4ML) + P) / 6]
Where 'O' represents optimistic scenario, 'ML' represent most likely scenario, 'P' represents pessimistic scenario

FMEA Risk Priority Number (RPN) = S * O * D
Where 'S' represents severity rating, 'O' represents occurrence rating, 'D' represents detection rating

Mean ($\overline{x}$): The sum of all values in a data set divided by the count of all values: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Median: The middle value that separates the higher half from the lower half of the data set

Mode: The most frequent value in the data set

Range: The numeric difference between the greatest and smallest values in a data set

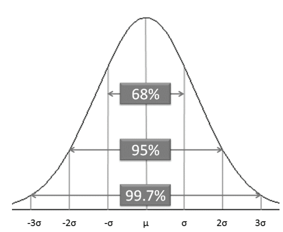
Variance (s^2): The average squared deviation of each value from its mean: $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Standard Deviation (s): The square root of the variance: $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

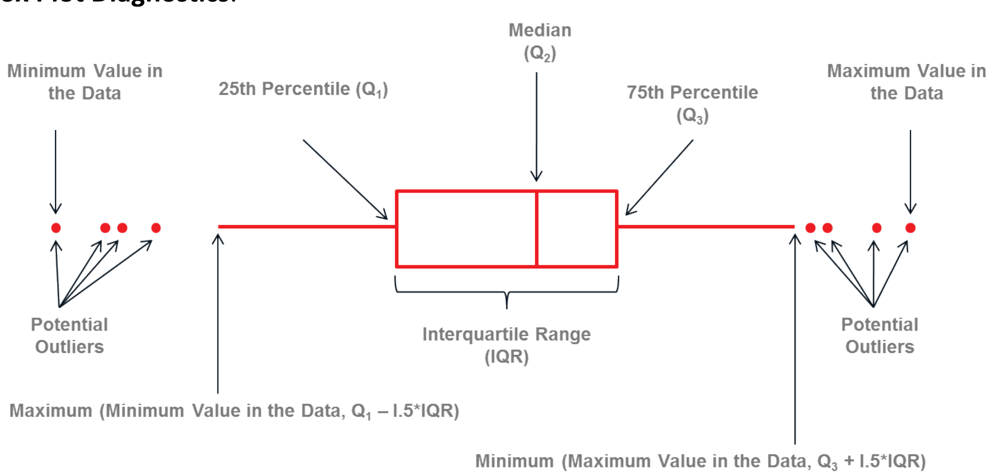
Z Score (z): Measures # of standard deviations an observation is above or below the mean: $z = \frac{x - \mu}{\sigma}$

68-95-99.7 Rule: The empirical rule in statistics states that for a normal distribution:

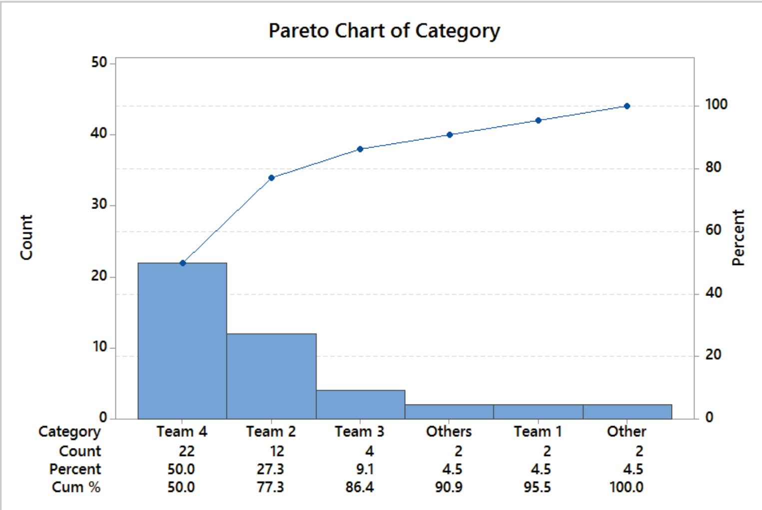
- *About 68% of the data fall within one standard deviation of the mean*
- *About 95% of the data fall within two standard deviations of the mean*
- *About 99.7% of the data fall within three standard deviations of the mean.*



Box Plot Diagnostics:



Pareto Principle: commonly known as the "law of the vital few" or "80:20 rule." It means that the majority (approximately 80%) of effects come from a few (approximately 20%) of the causes. The Pareto principle was named after Italian economist Vilfredo Pareto, who observed in 1906 that 80% of the land in Italy was owned by 20% of the population.



A ***Pareto chart*** is a chart of descending bars with an ascending cumulative line on the top.

Percent to Total: A Pareto chart shows a table with the percentage to the total for individual bar.

Cumulative Percentage: A Pareto chart also shows a table with the cumulative percentage of each additional bar. The data points of all cumulative percentages are connected into an ascending line on the top of all bars.

Measurement Systems Analysis (MSA): Key Components

%Contribution: The percent of contribution for a source is 100 times the variance component for that source divided by the total variation.

%Study Var (6*SD): The percent of study variation for a source is 100 times the study variation for that source divided by the total variation.

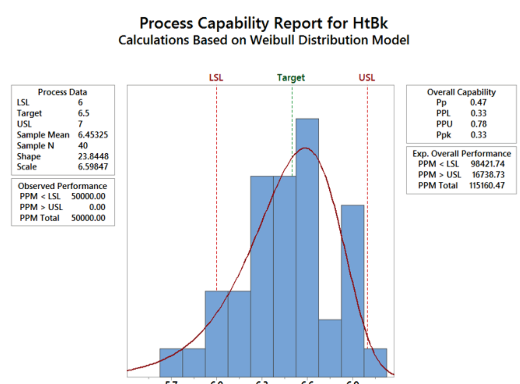
Distinct Categories: The number of distinct categories of parts that the measurement system is able to distinguish. If a measurement system is not capable of distinguishing at least five types of parts, it is probably not adequate.

Measurement System	% Study Var	% Contribution	Distinct Categories
Acceptable	10% or less	1% or Less	5 or Greater
Marginal	10% - 30%	1% - 9%	
Unacceptable	30% or Greater	9% or Greater	Less than 5

Process Capability: Measures how well a process performs relative to specification limits. Cp or Pp values greater than 1 indicate that process variation is less than the distance between the upper and lower specification limits. This is good, values greater than 1 are desirable. The higher Cp or Pp the better performing the process is. If Cp or Pp values are less than 1 it indicate that process variation is greater than the width of the USL and LSL, this is bad, there will be defects.

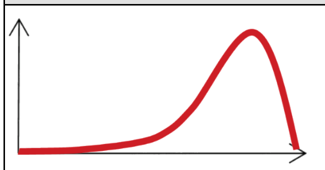
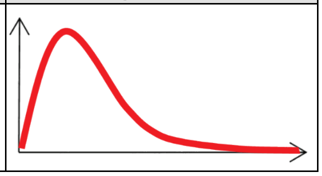
Cp	Pp	Cpm
$C_p = \dfrac{USL - LSL}{6 \times \sigma_{within}}$	$P_p = \dfrac{USL - LSL}{6 \times \sigma_{overall}}$	$C_{pm} = \dfrac{\min(T - LSL, USL - T)}{3 \times \sqrt{s^2 + (\mu - T)^2}}$

Cpk	Ppk
$C_{pk} = \min\left(\dfrac{USL - \mu}{3 \times \sigma_{within}}, \dfrac{\mu - LSL}{3 \times \sigma_{within}}\right)$	$P_{pk} = \min\left(\dfrac{USL - \mu}{3 \times \sigma_{overall}}, \dfrac{\mu - LSL}{3 \times \sigma_{overall}}\right)$

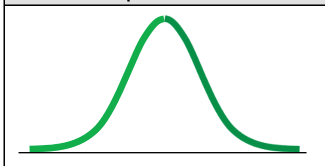
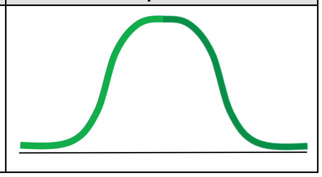


Where: USL & LSL are upper and lower specification limits
T is the specified target, μ is the mean

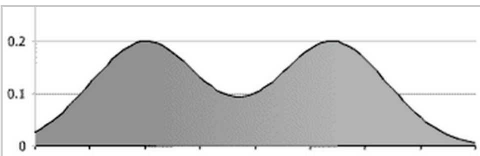
Skewness: Measure of the degree of asymmetry of the probability distribution

Left Skew	Right Skew
	

Kurtosis: Measure of the peakedness of the probability distribution

Leptokurtic	Platykurtic
	

Bi-Modal: A continuous probability distribution with two different unimodal distributions



Types of Distributions

Binomial distribution is a discrete probability distribution describing the probability of any outcome of the experiment.

Poisson distribution is a discrete probability distribution describing the probability of a number of events occurring at a known average rate and in a fixed period of time.

Normal distribution is a continuous probability distribution describing random variables which cluster around the mean.

Exponential distribution is a continuous probability distribution describing the probability of events occurring at a known constant average rate between points of time.

Weibull distribution is a continuous probability distribution which is widely used to model a great variety of data due to its flexibility (used in survival analysis, reliability/failure analysis).

Student's t distribution is a continuous probability distribution that resembles a normal distribution.

F distribution is a continuous probability distribution which arises in the analysis of variance or test of equality between two variances.

Chi-square distribution (also *chi-square* or *χ^2-distribution*) is a continuous probability distribution of the sum of squares of multiple independent standard normal random variables.

Population Parameters: An entire set of objects or observations

Mean	Standard Deviation	Variance	Median
μ	σ	σ^2	η

Sample Parameters: a subset of the population

Mean	Standard Deviation	Variance	Median
$\bar{x}$	S	S^2	$\tilde{x}$

Sample Size Calculations:

Continuous Data	Discrete Data
$n = \left(\frac{\left(Z_{\alpha/2} + Z_{\beta} \right) \times s}{d} \right)^2$	$n = \left(\frac{Z_{\alpha/2} + Z_{\beta}}{d} \right)^2 \times p \times (1-p)$

where:
***n** is the number of observations in the sample.*
***α** is the risk of committing a false positive error.*
***β** is the risk of committing a false negative error.*
***s** is the estimation of standard deviation in the population*
***d** is the size of effect you want to be able to detect.*
***p** is the proportion of one type of event occurring*

Sampling Strategies

Simple random samples are selected in such a way that each item in the population has an equal chance of being selected.

Matched random samples are samples randomly selected in pairs, each of which has the same attribute.

Stratified random samples are used when a population can be grouped or "stratified" into distinct and independent categories.

Systematic sampling has a random start and then every i^{th} item is selected going forward.

Cluster sampling is a sampling method in which samples are only selected from certain clusters or groups of the population.

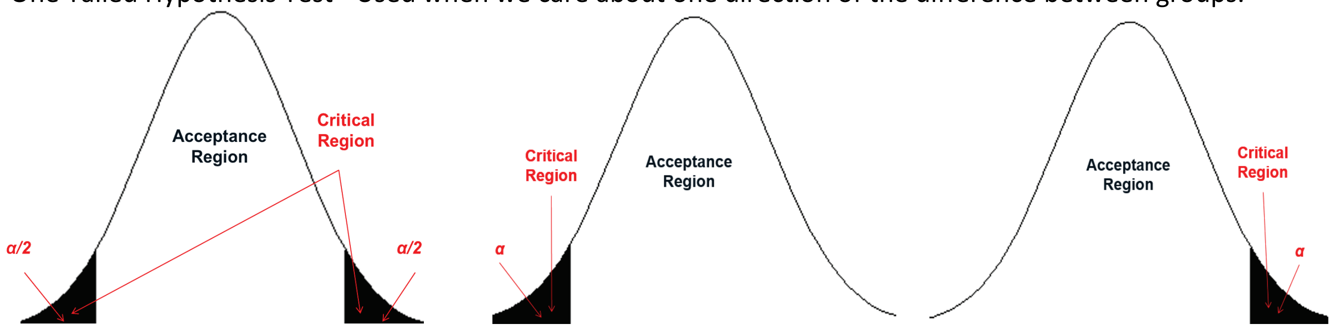
Hypothesis Testing: A statistical method used with a specific hypothesis about a population, the decision to reject the hypothesis is made based on the assessment of sample data.

- Null Hypothesis (H_0): no difference, no change, random chance
- Alternative Hypothesis (H_1 or H_a): difference exists, change occurred, observations affected by non-random chance

- Possible Conclusions of Hypothesis Testing:
  - Reject the Null *[Reject the Null Hypothesis when the p-value is less than 0.05 ("when the p is low the null must go")]*
  - Fail to Reject the Null
- Errors in Hypothesis Testing:
  - Type 1 Error (alpha error): Null hypothesis is rejected when it is true (rejected when you shouldn't have)
  - Type 2 Error (beta error): Null is not rejected when it is not true (didn't reject when you should have)

	Null hypothesis is true	Alternative hypothesis is true
Fail to reject null hypothesis	Correct	Incorrect (Type II Error)
Reject null hypothesis	Incorrect (Type I Error)	Correct

- Two Tailed Hypothesis Test - Used when we care about whether there is a difference between groups and we do not care about the direction of the difference.
- One Tailed Hypothesis Test - Used when we care about one direction of the difference between groups.



Correlation: A statistical technique that describes whether and how strongly two or more variables are related.
Pearson's correlation coefficient (r): Also called Pearson's r or coefficient of correlation
Correlation coefficients range from −1 to 1

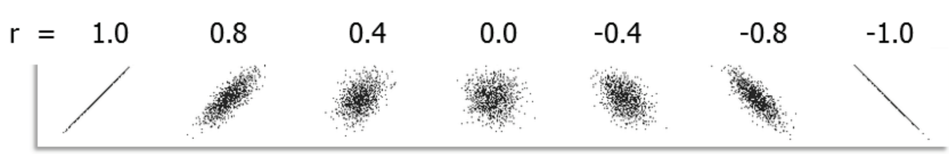
- If r = 0, there is no linear relationship between the variables

The *sign* of r indicates the *direction* of the relationship:

- If r < 0, there is a negative linear correlation.
- If r > 0, there is a positive linear correlation.

The *absolute value* of r describes the *strength* of the relationship:

- If |r| ≤ 0.5, there is a weak linear correlation.
- If |r| > 0.5, there is a strong linear correlation.
- If |r| = 1, there is a perfect linear correlation.



Multicollinearity: When two or more independent variables in a multiple regression model are correlated.
Variance Inflation Factor (VIF)*:* Quantifies the degree of multicollinearity for each independent variable in a model.
"Rules of Thumb" to analyze variance inflation factor (VIF):

- If VIF = 1, there is no multicollinearity.
- If 1 < VIF < 5, there is small multicollinearity.
- If VIF ≥ 5, there is medium multicollinearity.
- If VIF ≥ 10, there is large multicollinearity.

How to Deal with Multicollinearity

1. Increase the sample size.
2. Collect samples with a broader range for some predictors.
3. Remove the variable with high multicollinearity and high p-value.
4. Remove variables that are included more than once.
5. Combine correlated variables to create a new one.

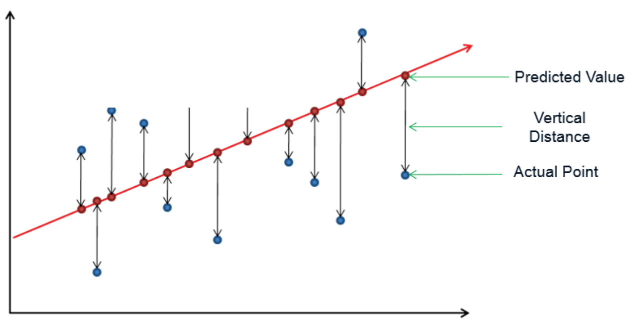
Confidence interval - The range in which the population mean of the dependent variable would fall with some certainty, given specified values of the independent variables.

The width of confidence interval is related to:

- Sample size
- Confidence level
- Variation in the data

Residuals: The vertical difference between actual values and predicted values or the "fitted line" created by a regression model. Residuals should be:

- Normally distributed with mean equal to zero
- Independent
- Constant variance
- The underlying population relationship is linear



DOE Key Terms
Response: Y, dependent variable, process output measurement
Effect: The change in the average response across two levels of a factor or between experimental conditions
Factor: X's, inputs, independent variables
Fixed Factor: Factor that can be controlled during the study
Random Factor: Factor that cannot be controlled during the study
Factor Levels: Factor settings, usually high and low, + and −, 1 and −1
Treatment Combination: setting of all factors to obtain one response measurement (also referred to as a "run")
Replication: Running the same treatment combination more than once (sequence of runs is typically randomized)
Repeat: Non-randomized replicate of all treatment combinations
Inference Space: Operating range of factors under study
Main Effect: The average change from one level setting to another for a single factor
Interaction: The combined effect of two factors independent of the main effect of each factor
Orthogonal: A DOE design is orthogonal if the main effect of one factor can be evaluated independently of other factors
Confounding: When two factors or factor interactions "confuse" our ability to know which factor has the effect on 'Y'
Resolution: The measure or degree of confounding. Higher resolution means less confounding
Full Factorial: (2^k) All possible combinations of factors and levels are created and tested; [2 levels, k factors (e.g. 2 levels, 3 factors = 8 runs, 2 levels 4 factors = 16 runs]
Fractional Factorial: (2^{k-1}) A subset of a full factorial experiment [2 levels, k factors but the number of runs is reduced (e.g. 2 levels, 3 factors but the experiment only has 2^{k-1} runs or 2^2 = 4 runs)]

7 Wastes (7 Muda)

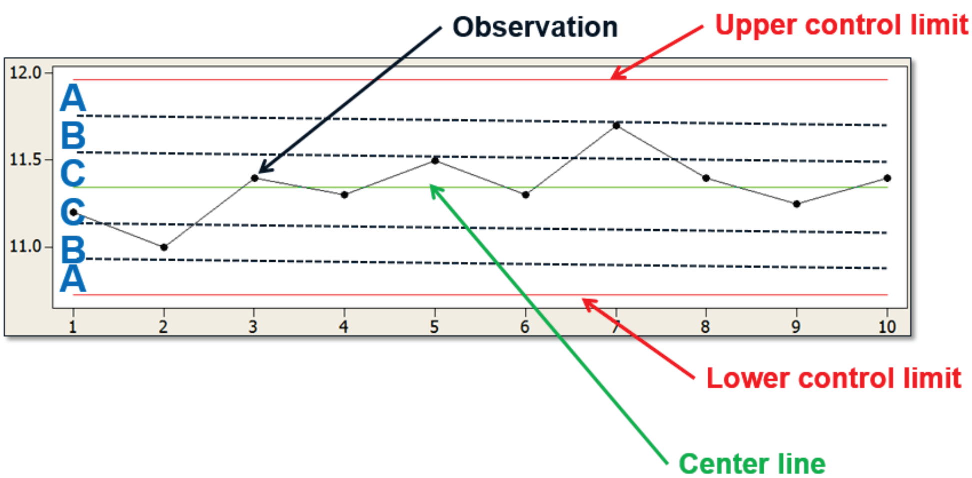
1. Defects
2. Overproduction
3. Over-Processing
4. Inventory
5. Motion
6. Transportation
7. Waiting

Five S (5S)

- *Seiri* (sorting)
- *Seiton* (straightening)
- *Seiso* (shining)
- *Seiketsu* (standardizing)
- *Shisuke* (sustaining)

Control Chart Composition:

- Zone A: between two and three standard deviations from the center line
- Zone B: between one and two standard deviations from the center line
- Zone C: within one standard deviation from the center line



Western Electric Rules for Statistical Process Control:

Test 1: One point more than three standard deviations from the center line (i.e., one point beyond zone A)
Test 2: Nine points in a row on the same side of the center line
Test 3: Six points in a row steadily increasing or steadily decreasing
Test 4: Fourteen points in a row alternating up and down
Test 5: Two out of three points in a row at least two standard deviations from the center line
Test 6: Four out of five points in a row at least one standard deviation from the center line
Test 7: Fifteen points in a row within one standard deviation from the center line
Test 8: Eight points in a row beyond one standard deviation from the center line

Return on Investment (ROI): $ROI = \frac{TotalNetBenefits}{TotalCosts} \times 100\%$

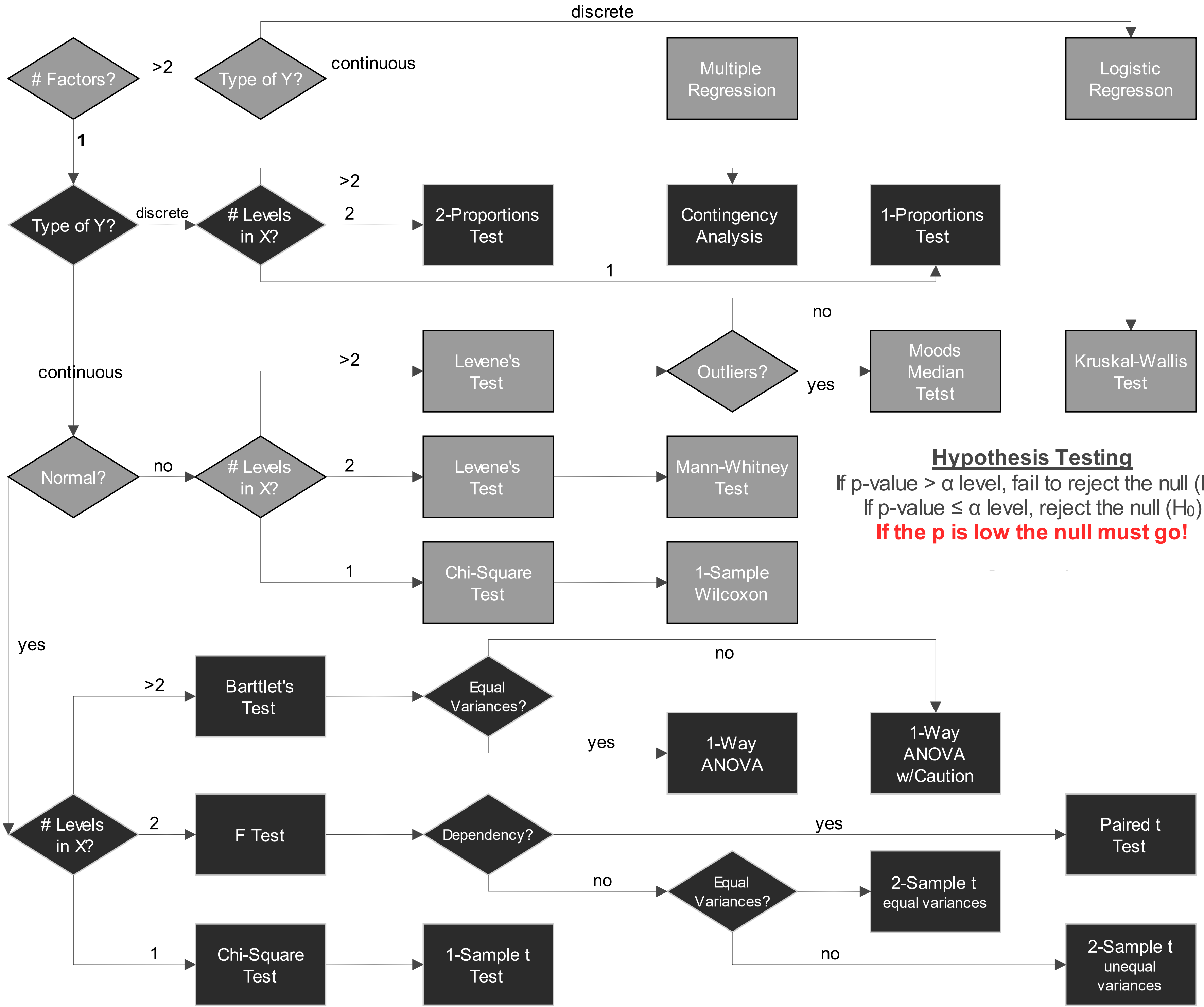
Net Present Value (NPV): $NPV = \frac{NetCashFlow_t}{(1+r)^t}$

$NetCashFlow_t$ is the net cash flow happening at time t
r is the discount rate
t is the time of the cash flow

NOTES

NOTES

Hypothesis Testing Roadmap

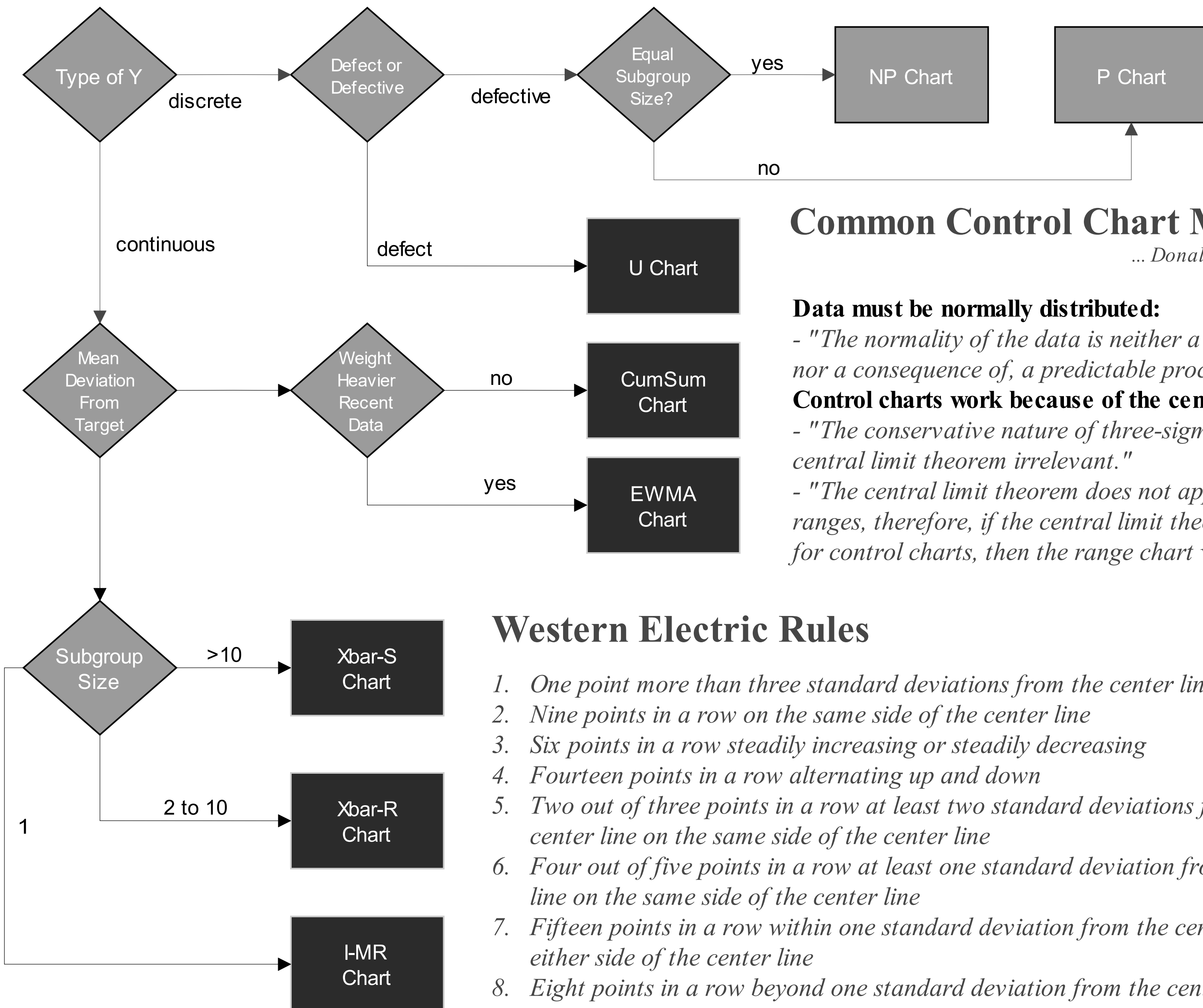


Hypothesis Testing

If p-value > α level, fail to reject the null (H_0)

If p-value ≤ α level, reject the null (H_0)

If the p is low the null must go!



Common Control Chart Myths

... Donald Wheeler

Data must be normally distributed:
- "The normality of the data is neither a prerequisite for, nor a consequence of, a predictable process."
Control charts work because of the central limit theorem:
- "The conservative nature of three-sigma limits makes the central limit theorem irrelevant."
- "The central limit theorem does not apply to subgroup ranges, therefore, if the central limit theorem were the basis for control charts, then the range chart would not work."

Western Electric Rules

1. *One point more than three standard deviations from the center line*
2. *Nine points in a row on the same side of the center line*
3. *Six points in a row steadily increasing or steadily decreasing*
4. *Fourteen points in a row alternating up and down*
5. *Two out of three points in a row at least two standard deviations from the center line on the same side of the center line*
6. *Four out of five points in a row at least one standard deviation from the center line on the same side of the center line*
7. *Fifteen points in a row within one standard deviation from the center line on either side of the center line*
8. *Eight points in a row beyond one standard deviation from the center line on either side of the center line*