

Data Analytics For Lean Six Sigma

By-

SURBHI VERMA

ROLL NO. 1018

HM Sec- C

Course Description

The course is available on Coursera and is offered by University of Amsterdam. This course introduces data analytics techniques that are typically useful within Lean Six Sigma improvement projects. It is a 5 week course and requires 16 hours to complete.

Course Objectives:

- To analyze and interpret data gathered within Lean Six Sigma Improvement projects.
- To select appropriate tool for data based testing.
- To emphasize on use of data analytics tools and interpretation of the outcome.
- To learn techniques that can be used in a broader setting apart from improvement projects.
- To illustrate all the tools using different examples from actual lean six sigma projects.

,

Course Contents

WEEK 1- Data and Lean Six Sigma

- **Introduction to Lean Six Sigma** Process improvement is important to organizations and Lean Six Sigma is a popular method to do this. Data is important when you want to improve your processes. Data can help you to diagnose the current processes or to investigate the effects of your improvement actions.
- **Data and DMAIC** The lean Six Sigma project consist of five DMAIC phases:

<i>Define</i>	Select project and establish objectives.
<i>Measure</i>	Make the problem quantifiable and measurable.
<i>Analyze</i>	Analyze the current situation and make a diagnosis.
<i>Improve</i>	Develop and implement improvement actions.
<i>Control</i>	Control the improved process performance and close the project.

- **CTQ-** Critical To Quality characteristics is the property of a product or a service that is relevant to your project objective. What we want to change determines our CTQ. In statistics and other fields, this variable is called your Y-variable, the dependent variable, or the outcome variable.

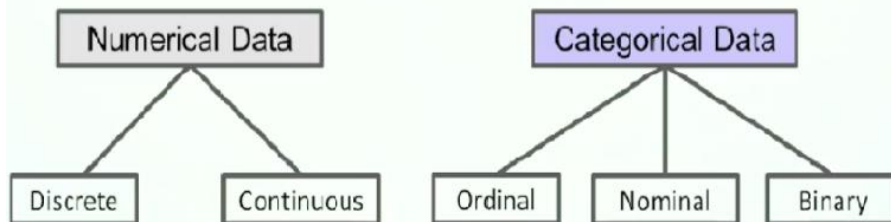
WEEK 1 Contd.....

- **CTQ-** Critical To Quality characteristics is the property of a product or a service that is relevant to your project objective. What we want to change determines our CTQ.
- **Units** are people, objects or phenomena you collect data on. It can be experimental units, observational units and measurement unit.
- **Sampling** It is important to select a representative sample. One way to obtain a representative sample is by randomly selecting the items.

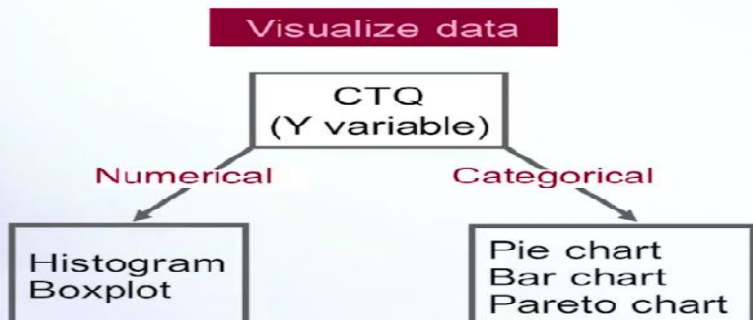
WEEK 2- Understanding and visualizing data

- **Numerical and Categorical data and Visualization** There are two types of data:
 - Qualitative data (Categorical)
 - Quantitative data (Numerical)

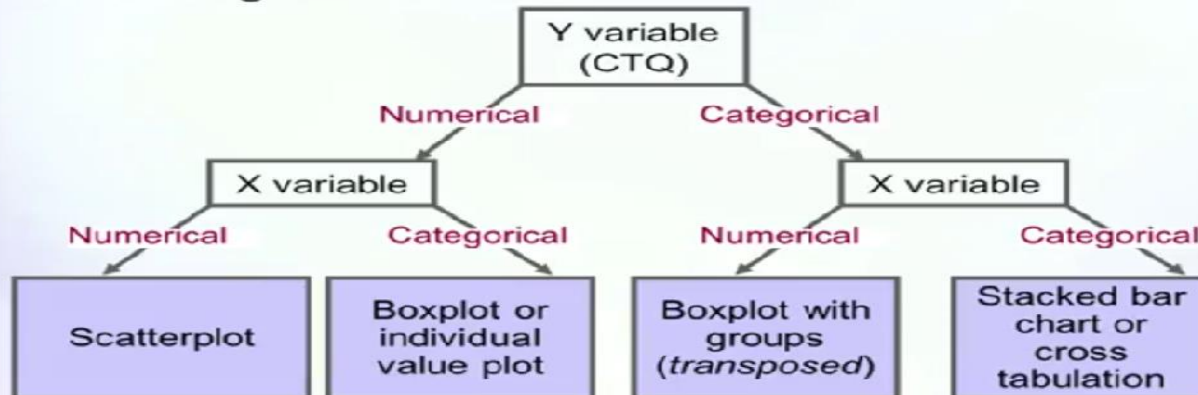
Numerical and categorical data



Numerical and categorical data



Visualizing two variables

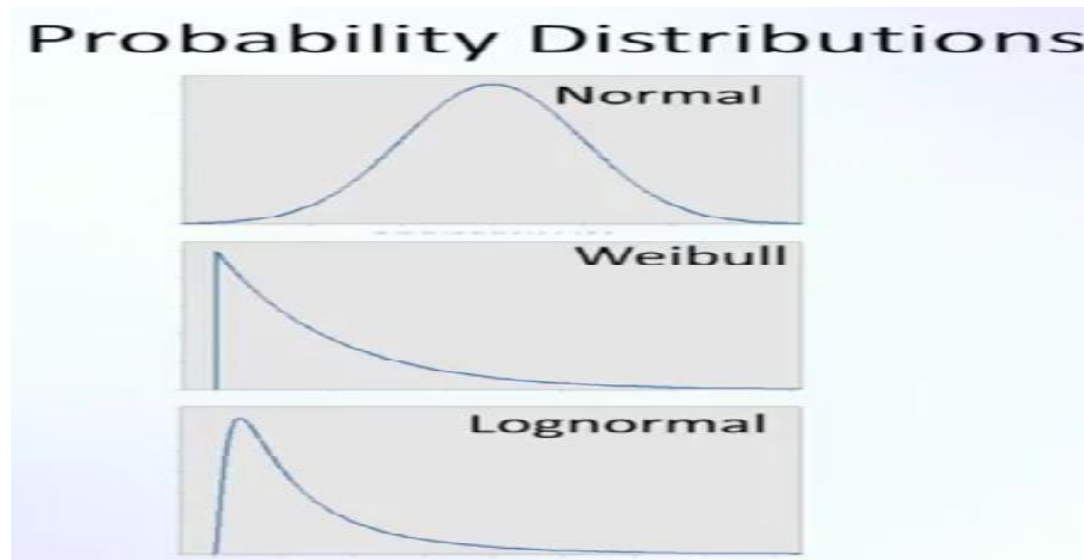


WEEK 3- Using probability distributions

- **Population versus Sampling**

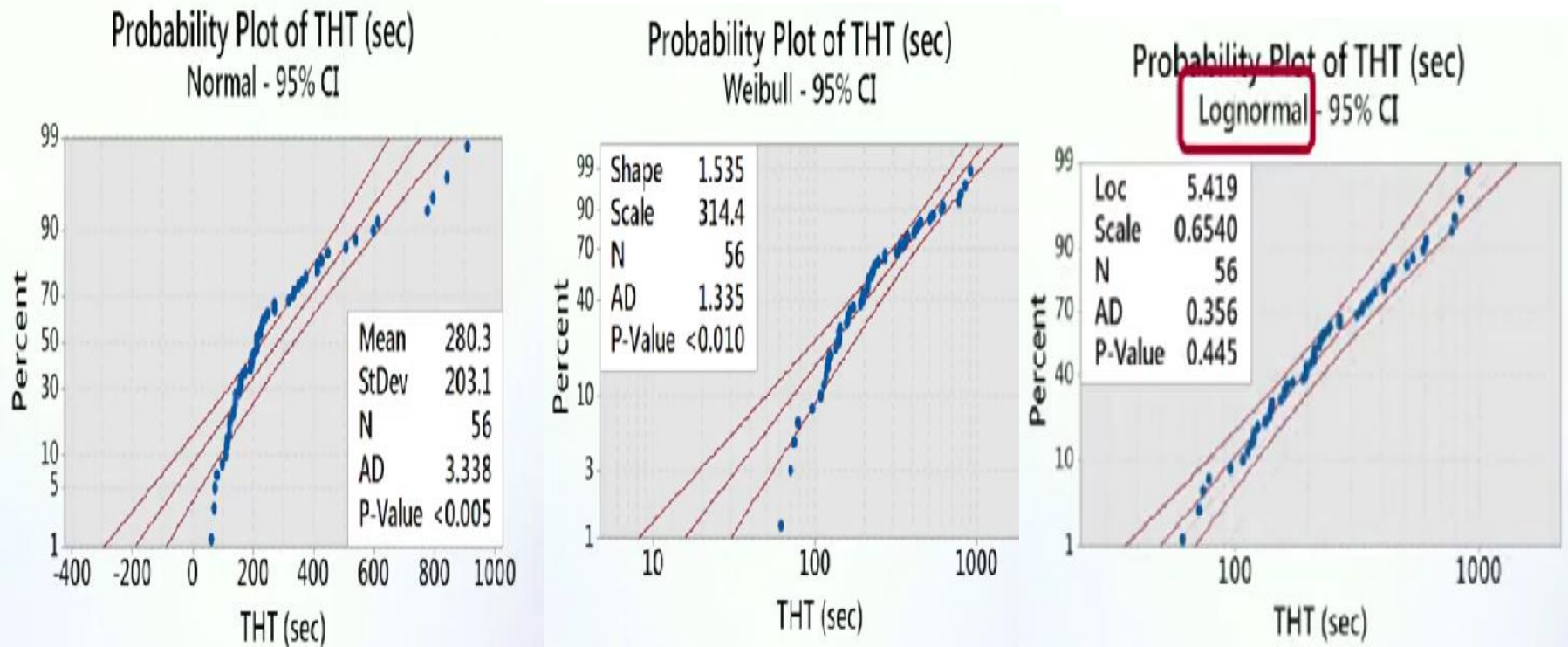
A population is a collection of all units of interest whereas a sample is a subset of the entire population, used to say something about the population.

Normal, Lognormal and Weibull distribution



WEEK 3 Contd.....

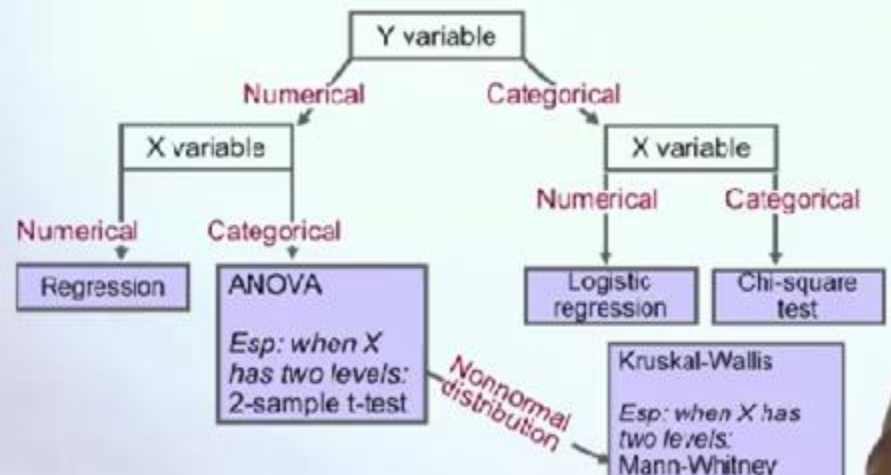
- **Probability plot** can be used to check if the distribution fits well if the measurements, the dots, are on a straight line. If we compare all three graphs, you can conclude that a lognormal distribution fits best to your data of THT. As a second check, you can compare the AD (Anderson Darling) values. The lower the value, the better the fit. In this case, the lognormal distribution has a lowest AD value and therefore fits the data best.



WEEK 4- a) Introduction to testing

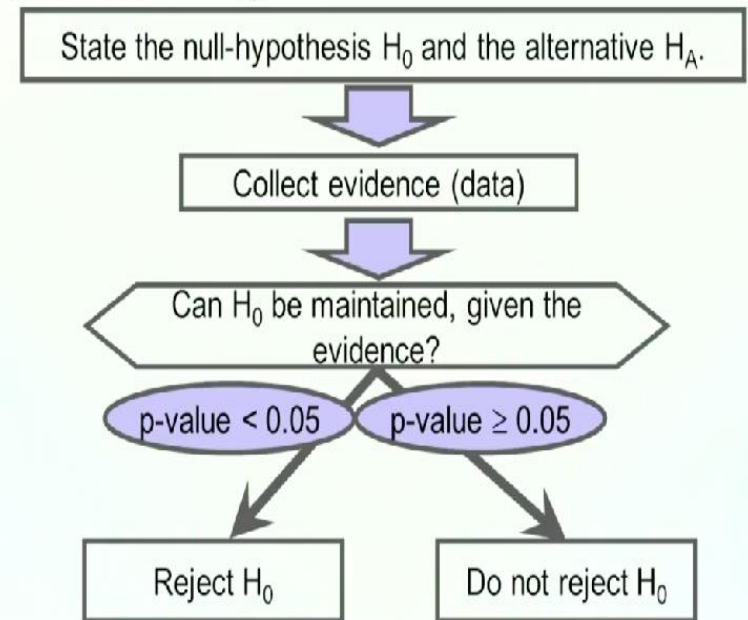
- Introduction to Data Analysis

Introduction to data analysis – *Summary*



- Hypothesis Testing

Hypothesis testing



WEEK 4- b) Testing: numerical Y and categorical X

- **ANOVA Analysis** ANOVA stands for Analysis of Variance . It is used when Y variable is numerical and the X variable is categorical. The **p-value** tells us whether the difference between the means of the machines are significant. The **R squared value** tells us how relevant the effect is.
- **Kruskal -Wallis Test** should be performed when assumptions of the ANOVA analysis are not met. To interpret the output, we have to look at the medians for each group and of course, add the P-value to see if this difference is statistically significant.
- **Two sample T-test** is used to compare means of two groups. The p-value can be used to calculate whether the differences of the two mean is coincidental, or truly significant.
- **Test for equal variances:** It shows that understanding variation is very important for the quality of a process. It helps to compare variances across groups and can be used to improve the p-value over and over.

WEEK 5- a) Testing: numerical Y and numerical Y

- **Correlation** is an index that expresses how strong a relationship between two numerical variables is. This index is always between minus 1 and plus 1
- **Regression analysis** is about finding the best fitting linear line for your data. It is a statistical method to test if two numerical variables are related by calculating the best fitting line.

b) Testing: Categorical Y

- **Chi-Square Analysis** The two steps of the chi-square analysis are:
 - to make a cross tabulation
 - p-value shows you whether or not the effect you found is significant
- **Logistic regression** It consists of three steps:
 - In the first step, you may need to convert your data to the correct format.
 - In the second step, you make a fitted line plot and evaluate the quality of the fit.
 - In the final step, you analyze the p-value and see if you have found a significant relationship.

KEY LEARNINGS

- **WEEK 1:** In this week we came to learn about the lean six sigma, Data and DMAIC framework, how to select CTQs, how to perform sampling, organizing the data, the units and operational definition along with installation of Minitab.
- **WEEK 2:** In this week we learn to visualize numerical and categorical data, descriptive statistics, Pareto analysis, visualizing two variables along with video exercises on investigation time and coffee batch
- **WEEK 3:** In this week we got to know the difference between population and sampling, we learn about estimation, empirical CDF and confidence intervals, normal, lognormal and weibull distribution curves, probability plot and a video exercise on the length of stay.
- **WEEK 4:** In this week we got to learn about data analysis, hypothesis testing, causality, ANOVA analysis, ANOVA residual analysis, Kruskal-Wallis test, Two sample t-test, test for equality of variances and video exercises on productivity and department.
- **WEEK 5:** In this week we learn about correlation, regression analysis, regression residual analysis, regression prediction interval, quadratic regression, chi-square analysis, logistic regression and video exercises on picking, printers and students.

**THANK
YOU!**

