Report on Online Course

By

Amit Kumar Thakur

MBA (Hospital and Health Management)
2019-21

ONLINE COURSE- DATA ANALYTICS FOR LEAN SIX SIGMA

HOURS

- Approximately 15 hours to complete.

LEARNING METHODOLOGY

- Video lectures, cases, Assignments, and Reading material.

WEBSITE/ ORGANIZATION

- https://www.coursera.org

COURSE DESCRIPTION

- Data analytics techniques that are typically useful within Lean Six Sigma improvement projects.
- To analyze and interpret data gathered within such a project.
- Use of Minitab to analyze the data.
- Learnt what Lean Six Sigma is.
- Learnt to use data analytics tools and interpretation of the outcome.
- Use of Statistics for data analysis.

OBJECTIVE

- Objective of this course is to learn the data analytics techniques that are typically useful within Lean Six Sigma improvement projects.

KEY LEARNINGS

- Statistics
- Lean Six Sigma
- Data Analysis
- Minitab

INTRODUCTION TO LEAN SIX SIGMA- SUMMARY

- Process involvement is important
- Lean Six Sigma is a popular method to improve processes
- Data is important in process improvement to:

A. Find biggest problem
B. Diagnose current process
C. Investigate effect of improvements

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

SELECTING CTQs (Critical to Quality)

A property of a process, product or service that is relevant to the project objective.

CTQ is a Lean Six Sigma terminology, in statistics this is Y-variable, dependent variable or outcome variable.

UNITS AND OPERATIONAL DEFINITION

- **Units**: the ‘people, objects or phenomena’ you collect data on
- **Measurement units**: There is never one correct operational definition. Instead, you should look for the most sensible definition.

SAMPLING

Sampling is done because: -

- It is difficult to get data on entire population
- It gives precise information about a part of the population, than vague information about the whole population

How do we collect a sample?

- The sample has to representative
- To this extent: the sample should be based on a mechanism that is independent of the question.

ORGANISING YOUR DATA.

A data set consists of

- UNITS: people, objects, or phenomena
- VARIABLES: properties that are measured

Organize your dataset as

- Units in the ROWS
- Variables in the COLUMNS

NUMERICAL AND OBSERVATIONAL DATA

NUMERICAL (QUANTITATIVE) DATA	CATEGORICAL (QUALITATIVE) DATA
• Discrete • Continuous	• Ordinal • Nominal • Binary

VISUALIZING DATA

If **Y- variable** of your CTQ is **Numerical**
You can make Histogram or boxplot for it.

Categorical
You can make Pie chart, Bar chart, Pareto chart.

Descriptive statistics

- Summarizing the **numerical data** into just a few key figures, and these key figures are called descriptive studies.
- Descriptive statistics are only available for **numerical data** only because they are based on mathematical operations like addition and multiplication.
- For categorical data, you can make a table

Histogram and boxplot

- When you have a numerical data, histogram and boxplot are the most popular methods for visualization of this variable.

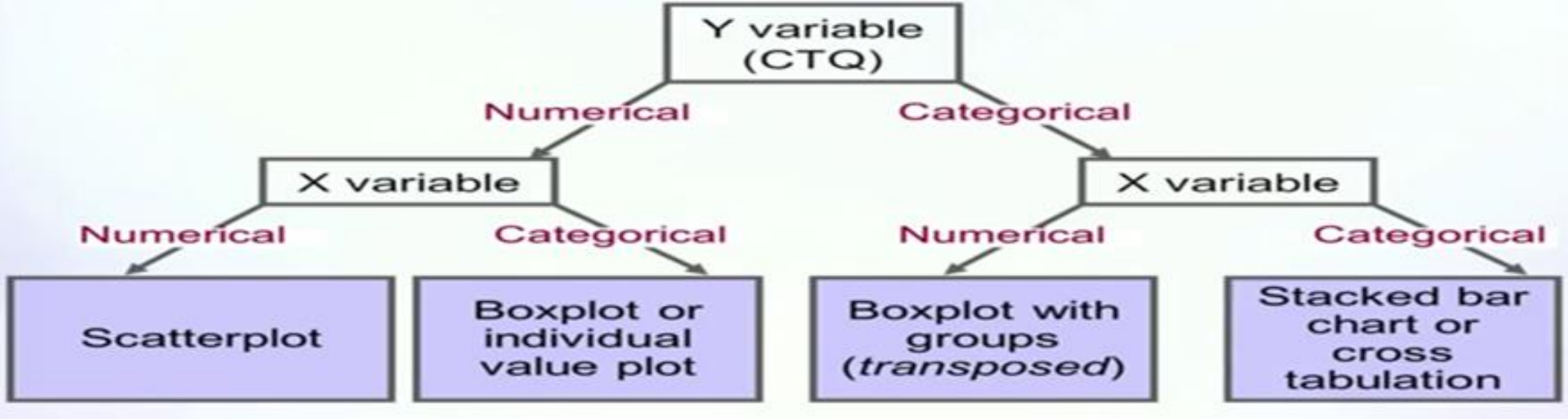
Visualizing categorical data

- A pie chart shows us how often each category occurs.
- Bar chart gives us the same information, but it is easier to spot which category occurs most often.
- A tally table gives the exact numbers

PARETO CHART (80/20 PRINCIPLE)

- 80% of all problems can be attributed to 20% of the causes. This is Pareto's law.
- It is known as 80/20 principle, or Pareto principal, and known as the rule of the vital few and the trivial many.
- Only a few issues cause most of the problems.
- A pareto chart can help you single out the few vital issues, from the trivial many

VISUALIZING TWO VARIABLES



ESTIMATION AND CONFIDENCE INTERVAL

- We use confidence interval to quantify this uncertainty.
- It gives you boundaries where the population mean will be in between, with a certain level of confidence.

PROBABILITY DISTRIBUTION

- There are hundreds of different distributions.
- Frequently occurring distributions are: -

1. Normal (bell shaped)
2. Weibull (skewed to right because tail is towards right)
3. Lognormal (also a skewed distribution curve)

EMPIRICAL CDF

- The empirical CDF is used to determine how many percent of the cases meet your service level agreement, or your specification.
- For instance, imagine you run a project in a call center, and you want to improve the total handling time. Your selected CTQ or your Y variable is the total handling time of THT. For the project you wonder, how often do my employees handle a call within 240 seconds, which is your target.
- You can use the crosshair function or the percentile lines to find these percentages. The Empirical CDF is a useful tool in the analyze phase of lean six sigma project.

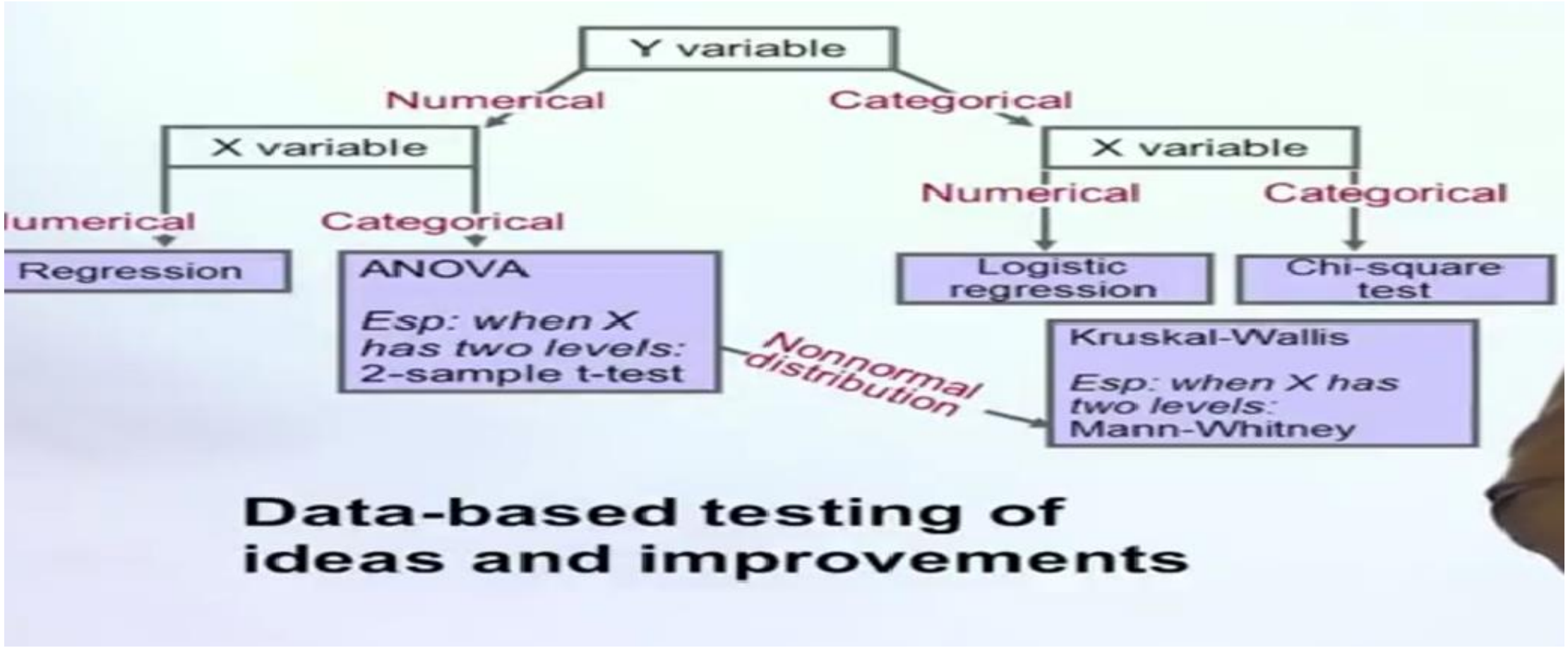
PROPERTIES OF NORMAL DISTRIBUTION

- It is symmetrical and it looks like a little bit like a bell. That is why it is also called the Bell-shaped distribution.
- Carl Friedrich Gauss is often associated with this distribution and its form. He made a significant contribution to the normal Distribution. That is why the Normal Distribution is also often called the Gauss curve.
- A normal distribution has two parameters that determine the shape of its distribution. The first parameter is the location parameter which determines the midpoint. This is denoted by mu, which is the location or mean.
- The second parameter describes the shape of the distribution and is known as the dispersion parameter, denoted by sigma. And it determines the spread of the distribution.

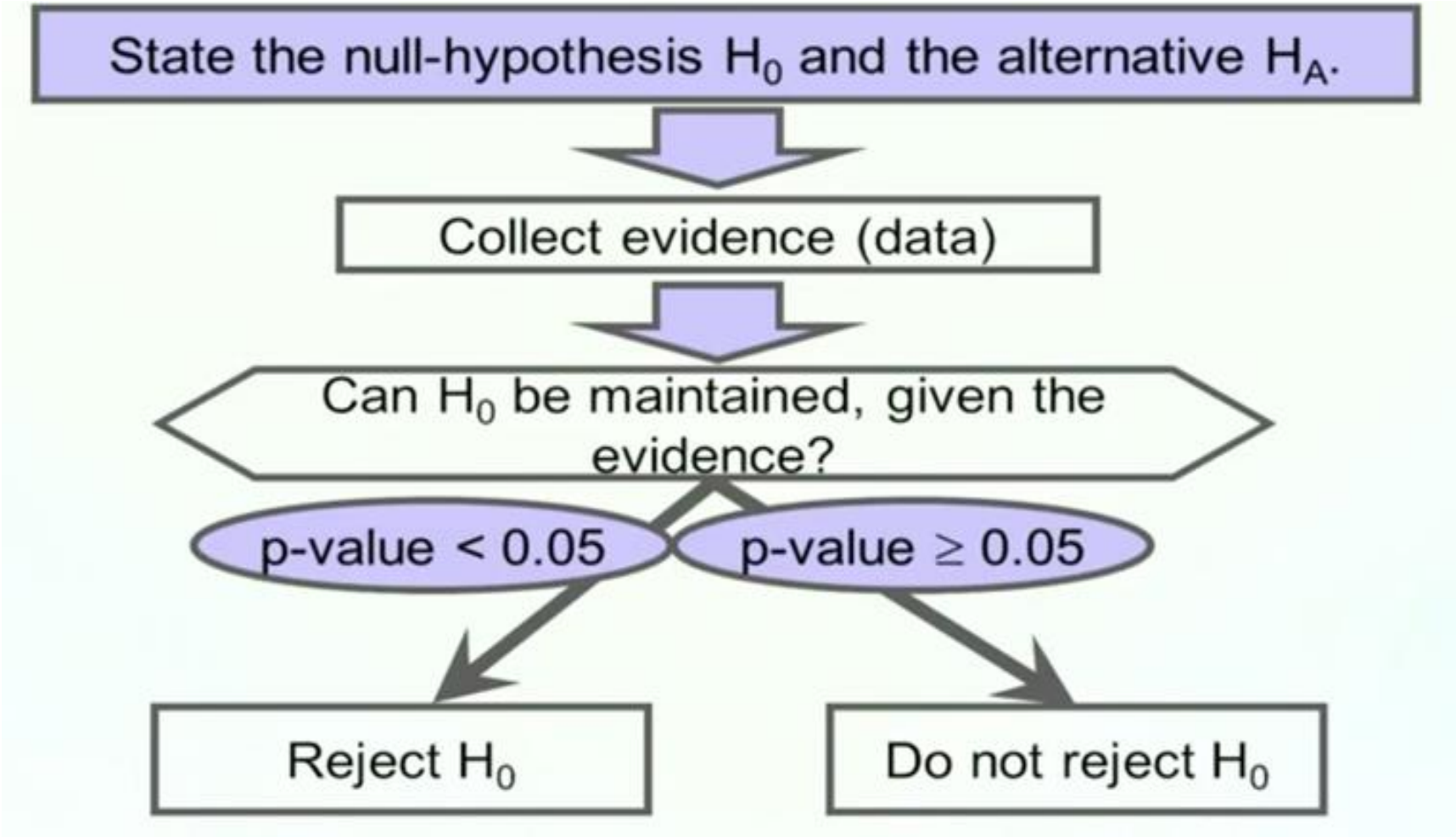
INTRODUCTION TO DATA ANALYSIS

- Data-based testing of ideas and improvements. This is one of the principles of Lean Six Sigma.
- This principle is sometimes referred to as evidence-based improvement actions.
- To analyze if two variables are related, you must determine which variable is your influence factor, or your X variable, and which variable is your CTQ, or your Y. Next, you must determine whether your variables are categorical or numerical.
- Once you have determined what type of data you have, the tree diagram can be a very useful tool to determine which type of analysis method is suitable for your data. And these are the methods that you can use for the data-based testing of ideas and improvements in your Lean Six Sigma project

Introduction to data analysis- Summary



HYPOTHESIS TESTING

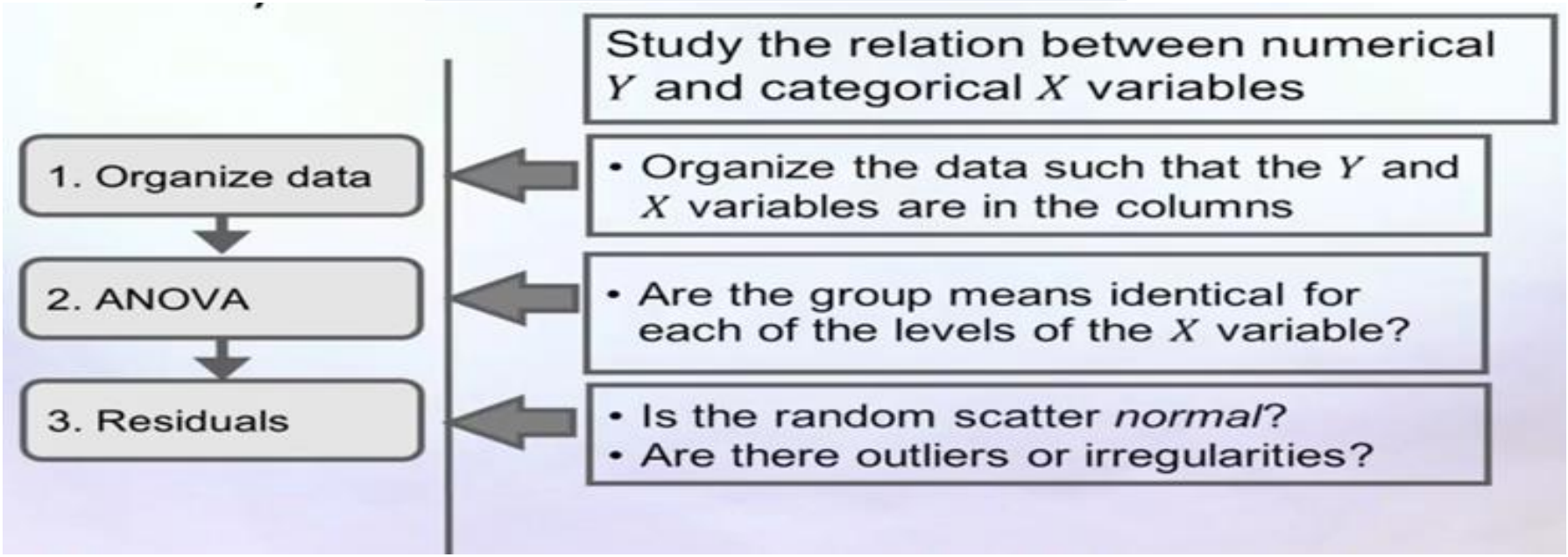


CAUSALITY

- **“Correlation does not imply causation”**

- It is about causes and effect relationship. This means when you change your x variable and nothing else your Y variable will change as a result. This is often very straight forward.

INTRODUCTION TO ANOVA



KRUSKAL-WALLIS TEST

- The Kruskal-wallis procedure is a good alternative for the ANOVA if the ANOVA assumptions are not met.
- To interpret the output, you must look at:
- The median for each group
- The p-value to determine if the results are significant

TWO SAMPLE T-TEST

- A two-sample t-test is a simple test to compare the means of two different groups.
- The p-value can be used to calculate whether the difference of the two means are coincidental or truly significant.
- The difference between one-way ANOVA and two sample t-test is that, one-way ANOVA is used to compare means of more than 2 groups at a single time.

TEST FOR EQUALITY OF VARIANCES

- Increasing variability always degrades the performance of a production system. This is the law of variability.

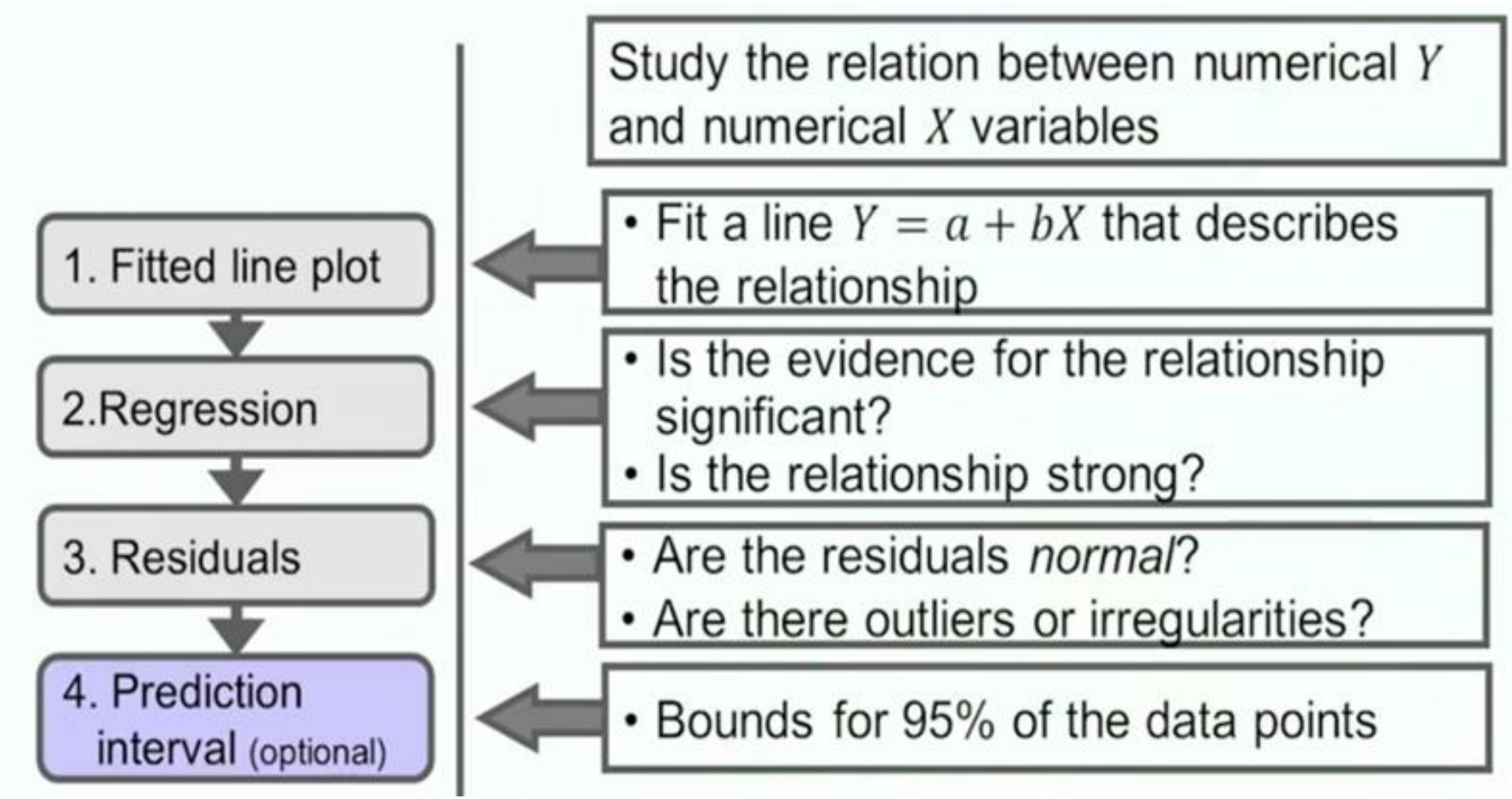
CORRELATION

- Correlation is an index that expresses how strong a relationship between two numerical variables is.
- The correlation index is always between -1 and 1
- 0 means no relationship
- Positive correlation means that if one variable increase then other variable also increases, while Negative correlation means that if x variable increases, then Y variable decreases.

INTRODUCTION TO REGRESSION

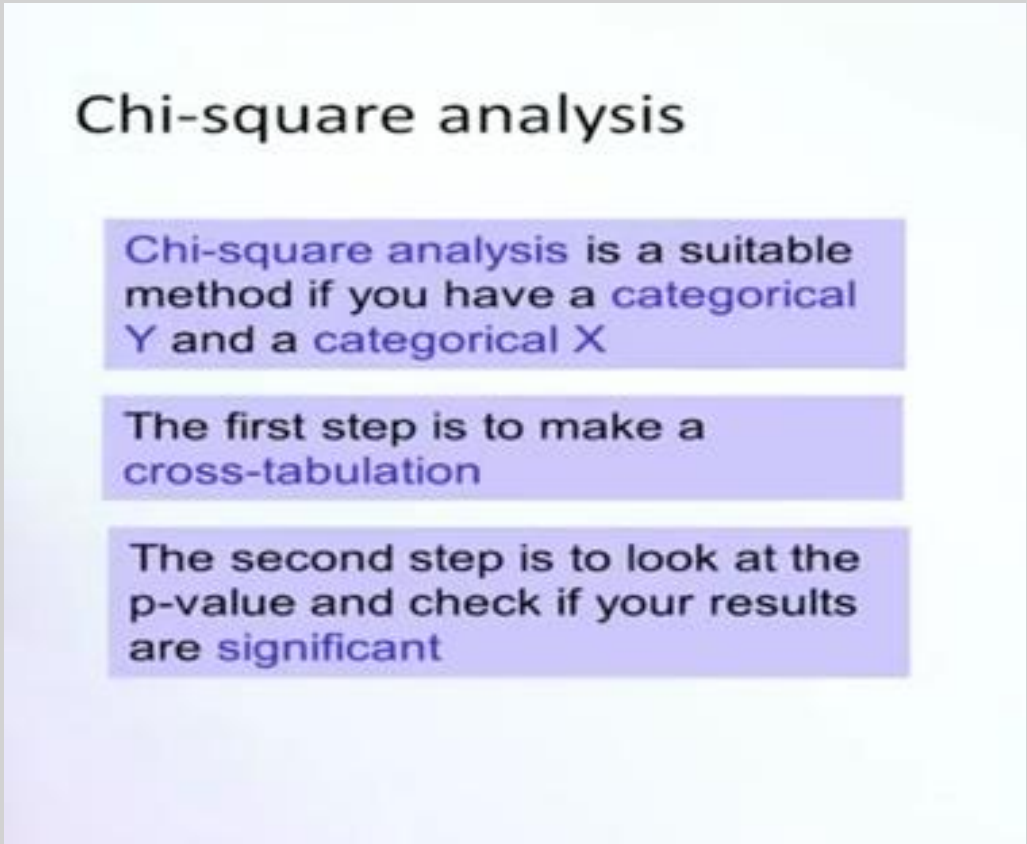
- Regression analysis is a statistical method to test if two numerical variables are related to each other
- Regression analysis consists of four steps:

Regression steps

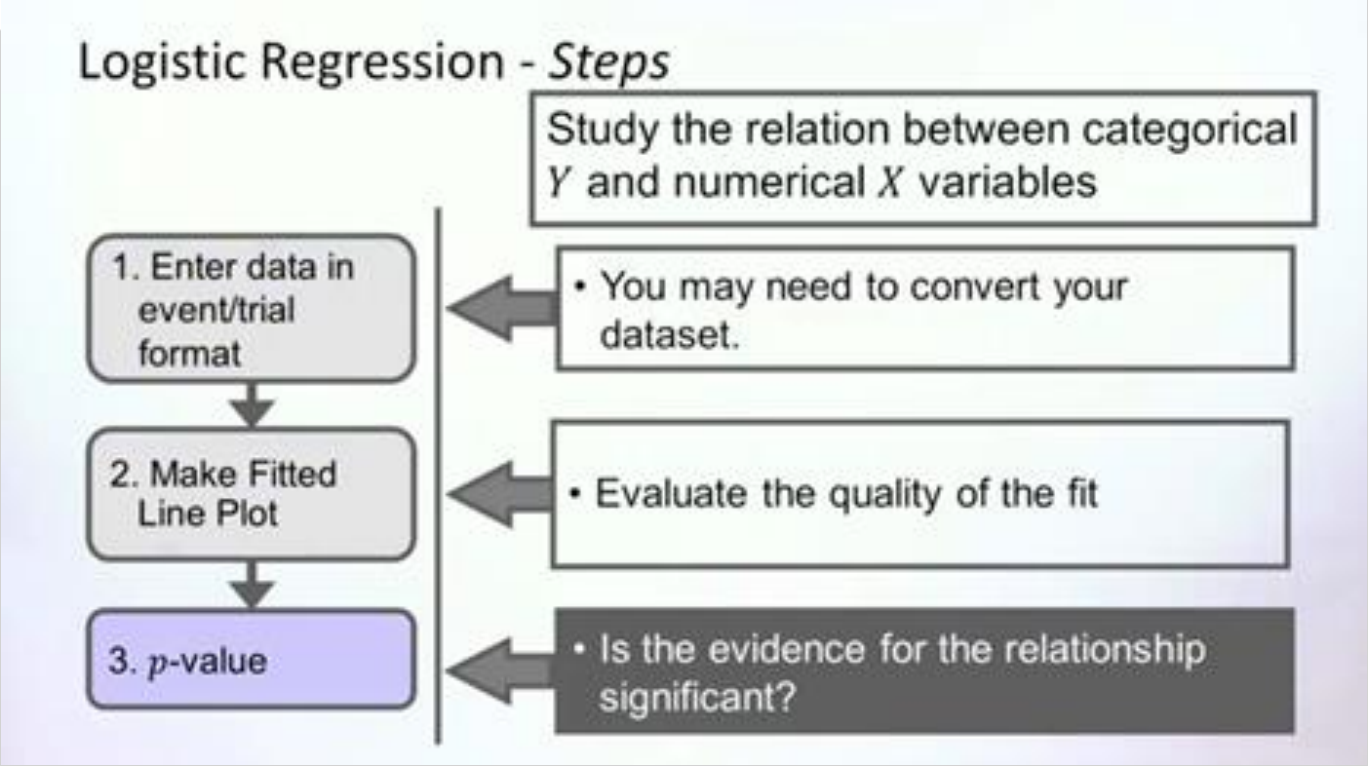


QUADRATIC REGRESSION

- If the assumptions of regression analysis are not met, you must perform an alternative technique.
- One of the alternatives is Quadratic Regression
- A quadratic regression fits a Curved line.



LOGISTIC REGRESSION ANALYSIS



Thank you