

SUMMER TRAINING REPORT

ON

REPORT 1: A STUDY ON THE IMPACT OF BEHAVIOUR INTERVENTIONS ON GENERAL POPULATION INTRODUCED BY GOVERNMENT OF INDIA AMID COVID-19 PANDEMIC.

REPORT 2: ONLINE COURSE REPORT ON “DATA ANALYTICS FOR LEAN SIX SIGMA”

BY

DR BHAVNA RANA

MBA HEALTH AND HOSPITAL MANAGEMENT

SESSION 2019-21



The IIHMR University, Jaipur

1, Prabhu Dayal Marg, Sanganer Airport
Jaipur-302029, Rajasthan
Ph: 0141 3924700, Fax: 3924738
Website: www.iihmr.edu.in

A study on the impact of Behaviour Interventions on general population introduced by Government of India
amid Covid-19 pandemic

By

Dr Bhavna Rana

MBA Hospital and Health Management

Academic Year 2019-21



The IIHMR University, Jaipur

1, Prabhu Dayal Marg, Sanganer Airport
Jaipur-302029, Rajasthan
Ph: 0141 3924700, Fax: 3924738
Website: www.iihmr.edu.in

ACKNOWLEDGEMENT

It has been a great privilege for me to be able to execute this study as a part of my curriculum as a post graduate student from Indian Institute of Health Management Research, Jaipur.

I would like to express my sincere and profound gratitude to my esteemed guide and mentor, Dr J.P. Singh, Professor, IIHMR University. He has always been there to guide us, despite of his busy schedule to discuss and guide us about the problems I was facing throughout the making of this report. His patience and perseverance gave me the confidence to be able to present my report in a very confident manner.

I would also like to thank my family and friends who have been a constant support for me during this study and pushed me to my maximum abilities. I am grateful to each one of you whole heartedly.

TABLE OF CONTENTS

1) List of Abbreviations.....	Page 4.
2) List of Tables.....	Page 4.
3) Introduction.....	Page 5.
4) Background.....	Page 5-7.
5) Research Question.....	Page 7.
6) Objectives.....	Page 7.
7) Methodology.....	Page 7-8.
Study design.....	Page 7.
Sampling	Page 7.
Data Collection.....	Page 8.
Participants.....	Page 8.
8) Results and Discussions.....	Page 9-16.
Knowledge.....	Page 10.
Practices.....	Page 11.
Attitudes and Opinions.....	Page 12.
9) Conclusions	Page 17-18.
10) References.....	Page 19.

ABBREVIATIONS

- 1) COVID-19 - Coronavirus Disease
- 2) GOI – Government of India
- 3) CDC – Centre for Disease Control
- 4) SARS-CoV2 – Severe Acute Respiratory Syndrome Coronavirus 2
- 5) PM – Prime Minister
- 6) UN – United Nations
- 7) MeitY – Ministry of Electronics and Information Technology
- 8) CAB – Citizenship Amendment Bill.
- 9) TV - Television

LIST OF TABLES

<u>S No.</u>	<u>Table</u>	<u>Page no.</u>
1.	Demographic variables: Age, Gender and no. of members in family	10.
2.	Demographic variables: State of belonging, occupation and educational status	11.
3	Correct knowledge about COVID-19 and interventions	12.
4	Practices followed by general public to control COVID-19 spread	13.
5	COVID-19 awareness message on phone calls	14.
6	TV advertisements on COVID-19 awareness	15.
7	Home quarantine as an approach against COVID-19	15.
8	Lockdown imposed by Government	16.
9	Calling back people stuck in other states/countries amid COVID-19 pandemic	16.
10	The most affected group due to lockdown in country	17.

INTRODUCTION

The rapidly growing Novel Coronavirus 2019 originated from the streets of Wuhan, China has placed an overwhelming burden on the health systems and authorities to respond with effective, immediate and appropriate interventions, policies as well as strategies to combat this life-threatening spread.

The most crucial step to reduce the transmission of the virus is rapid and widespread Behavioural Change. It has been observed that a perceived lack of consistency, competency, fairness, objectivity or sincerity in crisis response in the public may lead to fear and distrust.

Contrastingly, when the public perceives measures of having these characteristics and being informed and communicated through trusted and accessible channels and also being able to protect themselves. They are able to make informed choices when the necessary services are available helping them to comply with recommended practices in a better way.

Risk perceptions influence the protective behaviour of individuals but how people are perceiving the risk doesn't actually correlate with the actual risk. This was seen in the Influenza Pandemic during 2009-10 which had its outbreak in North America, where uncertainty and perceived exaggeration were also associated with reduced likelihood to implement the recommended behaviours.

Behaviour insights for Covid-19 is therefore, of a critical importance which includes knowledge about what drives the behaviour and awareness of changes among the people. Other psychological challenges are the mis information, stigmatisation or herd behaviour (such as hoarding of food and other essential commodities/panic buying).

BACKGROUND

Over more than 9 lakh cases and over 24,315 deaths from COVID-19 have been recorded in the country as on 14 July 2020. A major route of transmission of SARS-CoV2 is via droplet spread which requires close contact. In an attempt to mitigate the COVID-19 pandemic, many countries have adopted various interventions like physical distancing, compulsorily wearing of masks at public places, use of sanitizers, shut down of schools, colleges and complete lockdown to control the rising toll of COVID-19 cases.

India observed a 14 hour of voluntary public curfew (Janta Curfew) on the directions of PM Narendra Modi. It was a compulsory lockdown to be followed in the COVID-19 contaminated regions (hotspots) and all other major cities. Further, the lockdown for 21 days was again declared by PM on March 24th that was observed in the country. The PM again extended the nationwide lockdown till 3rd May which was to be followed by two-week extension as declared on 14 April. India's response to the pandemic was praised by the United Nations (UN) and WHO as "Comprehensive and robust", where the lockdown restrictions were termed as aggressive but were equally vital for containing the spread and building the necessary healthcare infrastructure.

GOVERNMENT RESPONSE AND INTERVENTIONS

- From 21st January, India began the thermal screening on the airports for the passengers arriving from China. It was carried out at 7 airports initially and then it was expanded to 20 airports to the end of January.
- The thermal screening was started for the passengers arriving from Thailand, Singapore, Hong Kong, Japan and South Korea during February.
- The other countries that got added to the list for thermal screening were Nepal, Vietnam, Indonesia and Malaysia by the end of February.
- On 17 March, the Government of India issued an advisory, urging all the Indian states to take social distancing measures as a preventive strategy for implementation till 31 March.
- The national and state helpline numbers were set up by Union as well as State Governments.
- Wearing of facial masks was made compulsory across major Indian cities and many states, where defaulters were made to pay the fine for not abiding this rule.
- States have also been asked to screen the people, quarantine them and to do periodic health check-ups.
- ZONAL CLASSIFICATION: The country was divided into the three zones – Green Zone, Red Zone, Orange Zone by the Government. Relaxations were made for the people living in those zones, accordingly.
 - 1) **Red Zone (Hotspots)**- These were the regions of high number of active cases and which had high doubling rate.
 - 2) **Orange Zone (Non-hotspots)**- These were districts that had fewer cases.
 - 3) **Green Zone** – Districts where there were no confirmed cases or without any new case of COVID-19 in last 21 days.
- AAROGYA SETU: A smart phone application called "Aarogya Setu" was launched by MeitY (Ministry of Electronics and Information Technology) to help in "contact tracing of the existing cases and containing the spread of COVID-19 pandemic in the nation.

RESEARCH QUESTION

- 1) What are the different behaviour interventions on general population introduced by Government of India amid COVID-19 pandemic?
- 2) What is the impact of different behaviour interventions on general population introduced by Government of India amid COVID-19 pandemic?

OBJECTIVES

- To identify different behaviour interventions introduced by Government of India amid COVID-19 pandemic.
- To identify different response patterns towards the interventions applied by Government of India.

METHODOLOGY

Study Design

This cross-sectional observational study included participants of age group 15-59 years from almost all the states of the country between 20 May and 15 June, 2020. After obtaining consent of each participant, a semi structured questionnaire was designed using Google forms (<https://docs.google.com/forms/>).

Sampling

Purposive sampling technique was applied to collect data from 384 respondents and taking into consideration of 10% non-response. The population of India is 1,210,854,977 as per Census 2011, out of which 756,784,361 (62.5%) is the population of age group 15-59 years that is economically active. Sample size is calculated by taking 95% confidence interval at 5% margin error.

Data Collection

The questionnaire was developed using questions on the basic interventions done by GOI where respondents were asked about the methods they were using to protect themselves from COVID-19 and what measures they have been taking at public places, their satisfaction level of the policies implemented by Government against COVID-19.

The questionnaire consisted of four sections namely demographic background (six questions), knowledge (three questions), practices (two questions), attitude and opinions (six questions). The link to the questionnaire was emailed and shared to the respondents using social networks like Facebook, whatsapp and Instagram. They received a system automated mail which clearly mentioned that before starting the survey, the participants could take up 20-25 mins to record their responses to the survey questions, in order to capture their response towards the behaviour interventions used by GOI amid COVID-19 pandemic. After reading a brief introduction of the study and providing informed consent, participants responded to the set of questions that appeared sequentially on the screen.

Participants

This cross-sectional study was conducted over a span of two weeks from 16 May to 30 May, 2020 during the lockdown period only. Data was collected online through a survey form in a specified time-line. People of the age group 15-59 years were specifically selected to be the part of this survey as of the reason that they are the economically active age group. A self-made questionnaire was circulated online which included which included brief introduction on the background, objective, voluntary nature of participation and declaration of confidentiality.

RESULTS AND DISCUSSIONS

The study had 384 respondents, out of which 285 responses were eligible for analysis. The majority of the respondents were females (n= 164, 57.5%) and most of the respondents were of the age group of 15-24 years (n= 149, 52.3%). Maximum number of the respondents were graduates according to their educational status (n= 128, 44.6%) and with a very slight difference were the post-graduates (n= 121, 42.5%). Most of the respondents were working in Government sector (41.4%), the remaining were into Private sector (18.9%), teachers (11.2%), students (11%), home makers (6%), doctors (8%), self employed (3.5%). The maximum number of respondents were from state Himachal Pradesh (32.4%) and the remaining were from Punjab(19%), Haryana (11%), Uttar Pradesh(8%), Rajasthan(8%), Maharashtra(5%), West Bengal(3.5%), Tamil Nadu(2.1%), Bihar (5%) and others (6%). The maximum respondents belonged to the family of four (34.4%), the respondents living in a family of less than four were few (13.7%) and the remaining belonged to the family of 5 (27.4%) and more than 5(24.6%). The demographic profile of the sample is shown in Table 1 and Table 2 below.

Table 1: Demographic characteristics of participants of age group 15-59 years from almost all the states of the country. (n =)

<u>CHARACTERISTIC</u>	<u>SUBGROUPS</u>	<u>N(%)</u>
AGE	15-24 years	149(52.3%)
	25-34 years	89(31.2%)
	35-44 years	32(11.2%)
	45-59 years	15(5.3%)
GENDER	MALE	121(42.5%)
	FEMALE	164(57.5%)
NO. OF MEMBERS IN FAMILY	LESS THAN 4	39(13.7%)
	4	98(34.4%)
	5	78(27.4%)
	MORE THAN 5	70(24.6%)

Table 2: Demographic characteristics of participants of age group 15-59 years from almost all the states of the country (n=.....)

<u>CHARACTERISTIC</u>	<u>SUBGROUPS</u>	<u>N(%)</u>
STATE OF BELONGING	HIMACHAL PRADESH	136(32.4%)
	PUNJAB	40(14%)
	HARYANA	13(4.6%)
	UTTAR PRADESH	38(13.3%)
	RAJASTHAN	11(3.9%)
	MAHARASHTRA	3(1.1%)
	JAMMU & KASHMIR	4(1.4%)
	DELHI	13(4.6%)
	OTHERS	27(9.5%)
OCCUPATION	GOVERNMENT SECTOR	17(6%)
	PRIVATE COMPANY JOBS	54(18.9%)
	TEACHERS	32(11.2%)
	DOCTORS	10(3.5%)
	STUDENTS	118(41.4%)
	HOME MAKERS	13(4.6%)
EDUCATION STATUS	POST-GRADUATES	121(42.5%)
	GRADUATES	127(44.6%)
	HIGH SCHOOL	37(13%)

KNOWLEDGE

- Almost all participants (96.5%) could specify the reason behind the lockdown which was imposed by GOI due to the spread of COVID-19.
- Almost all participants (91.2%) believed that the use of face masks and gloves was very helpful in order to deal with the current situation.
- More than 50% of the respondents knew about facemasks and had used face masks in their lives before the outbreak of COVID-19 also.

Table 3: Correct knowledge of COVID-19 and interventions used.

INDICATOR	N(%)
1) Reason behind the lockdown imposed by GOI	
-CAB law riots	1(0.4%)
-Spread of COVID-19	275(96.5%)
-Spread of Hay Fever	0(0%)
-None of the above	9(3.2%)
2) If use of facemasks and gloves was helpful in the situation	
-Yes	260(91.2%)
-No	5(1.8%)
-Maybe	20(7%)
3) If they knew about and used facemasks in their life before COVID-19 outbreak	
-Yes	159(56.6%)
-Never	56(19.9%)
-Maybe, don't remember	26(9.3%)
-Sometime	1(0.4%)

PRACTICES

The interventions that Government has advised the public to follow to contain the spread of COVID-19 and how people were taking those in practice were analysed here. More than three-fourth respondents(88.1%) used to follow Physical Distancing while going out to buy essential commodities where as some respondents(10.5%) accepted that they followed the physical distancing only sometime which showed a non serious attitude of public towards this practice where as only 1.1% denied of following physical distancing at public places. Almost all the respondents (99.6%) agreed that they wash hands every time when they come from outside.

Table 4: Practices being followed by public against COVID-19.

INDICATOR	N(%)
Following physical distancing while going out to buy essential commodities.	
- Always	251(88.1%)
- Sometimes	30(10.5%)
- Never	3(1.1%)
Washing hands after coming from outside.	
- Yes	284(99.6%)
- No	0%
- Sometimes	1(0.4%)

ATTITUDE AND OPINIONS

The respondents were asked about their opinions in the form of yes, no or may be about the interventions that GOI had introduced and been implementing lately, against COVID-19. This part helped to analyse the level of impact of these interventions on the general population including their satisfaction levels.

On being asked about their opinion on the awareness message being given by Government on the phone calls was a right approach or not, about three-fourth respondents (75.8%) believed that yes, it was a right approach to spread awareness against COVID-19 whereas some respondents (18.6%) believed it may be a right approach and few respondents (4.9%) said No to this approach. About 2 respondents (0.7%) had no clue about this type of approach being used by the Government. Refer Table 5.

Table5: COVID-19 Awareness message given on phone calls.

INDICATOR	N(%)
Awareness message on phone calls about COVID-19 is a right approach.	
Yes	216(75.8%)
No	14(4.9%)

Maybe	53(18.6%)
Don't know about this.	2(0.7%)

About less than three fourth respondents (71.9%) believed that the advertisements being commercialised on television was giving enough awareness to the general public which was required , whereas some respondents (14.4%) were not sure if the Tv advertisements were giving proper awareness where as few respondents (11.6%) were not satisfied with the level of information and awareness being provided was enough and about 4 respondents (1.4%) did not know about this. Refer Table 6.

Table 6: TV Advertisements on COVID-19

INDICATOR	N(%)
TV advertisements giving enough awareness about COVID-19 which is required.	
Yes	205(71.9%)
No	33(11.6%)
Not sure	41(14.4%)
Don't know about this	6(2.1%)

A larger part of 285 respondents (84.9%) agreed that the home quarantine that is advised by the Government is a right approach to save themselves from the COVID-19 pandemic whereas few respondents (8.8%) disagreed that home quarantine was a right approach to save themselves. Very few respondents (4.9%) were not sure if home quarantine was a right approach whereas about 4 respondents (1.4%) did not know about this approach. Refer Table 7.

Table 7: Home-quarantine as an approach towards COVID-19

INDICATOR	N(%)
Home-quarantine is a right approach advised to be followed to save oneself from being infected from COVID-19.	
Yes	242(84.9%)
No	25(8.8%)
Not Sure	14(4.9%)
Don't know	4(1.4%)

Majority of respondents (88.7%) believed that the Government's decision of imposing lockdown was helpful to contain the spread of COVID-19 whereas few respondents (9.2%) believed it maybe a right decision to impose lockdown by Government in the current situation. Only about 1.8% respondents were not satisfied with the decision of lockdown and only 1 respondent (0.4%) did not know about this. Refer Table 8.

Table 8: Lockdown imposed by Government

INDICATOR	N(%)
Government decision of imposing "Lockdown" is helpful.	
Yes	251(88.7%)
No	5(1.8%)
Maybe	26(9.2%)
Don't know about this	1(0.4%)

On being asked about the most vulnerable and affected group due to the lockdown imposed by GOI, more than three fourth respondents (77.2%) believed that Labourers were worst affected by this lockdown, where as few believed that Students (6%), employees (6%), senior citizens (4.9%), small businessmen (2.8%), small children (1.8%) and home makers (1.4%) were also affected due to this lockdown. Refer Table 9.

Table 9: Calling back people stuck in other states/countries amid COVID-19 pandemic.

INDICATOR	N(%)
Central/State Government(s) putting enough efforts to call back their people stuck in other states/countries.	
Yes	160(56.1%)
No	52(18.2%)
Not Sure	66(23.2%)
Don't know about this.	7(2.5%)

More than half of the respondents (56.1%) believed that the state/central Government(s) were putting enough efforts to call back their people (labourers, students etc) who were stuck in other states/countries whereas some respondents (23.2%) were still not sure about this. Few respondents (18.2%) believed that the efforts made by the Government were not enough in that case. About 7 respondents (2.5%) did not know about this. Refer Table 10.

Table 10: The most affected group due to lockdown amid COVID-19

INDICATOR	N(%)
The most affected group due to lockdown amid COVID-19 pandemic.	
Labourers	220(77.2%)
Students	17(6%)
Employees	17(6%)
Senior citizens	14(4.9%)

Small businessmen	8(2.8%)
Small children	5(1.8%)
Homemakers	4(1.4%)

CONCLUSIONS

A SWOT analysis of the various interventions introduced by the Government shows as following:

Strengths:

- Government took immediate measures to control the spread of COVID-19 by imposing travel restrictions, lockdown and introducing various other interventions to create awareness among the general population.
- Thermal screening was made mandatory at airports, offices and all the public places.
- Developing strategies to aware people and help them to be able to assess themselves through an application like Aarogya Setu, where people can self-assess themselves without giving burden to health care institutions.
- As per the survey conducted, maximum responses were in the favour of the interventions introduced by Government, most of the respondents were following the basic approaches like physical distancing, using face masks and gloves and washing hands.

Weaknesses:

- Even though Government adopted various strategies to control the spread of COVID-19, the toll of the infected is still rising exponentially where there is a need to control the spread by educating people more about it.
- There still is lack of knowledge and awareness in some parts of the country where people still don't have proper knowledge about the implementations drawn by Government and don't know the reasons behind them.
- The communication by Government is still unclear that risks in creating confusions among people.
- Government still need to adopt policies of taking out the people who are stuck in other regions of country and across the Globe.

Opportunities:

- The Government can do something for increasing the spread of awareness through campaigns and various other channels.
- Can take up use of R&D ecosystem for rapid development of COVID-19 antidote.
- Development of a strong-tier governance with stronger policies that should reach every single person in the rural areas of the country.
- Making use of more communication channels like Television, radio, internet etc.

Threats:

- Failure of the contact tracing leads to the worsening of the current situation of the pandemic.
- People still showing non-serious attitude towards lockdown and physical distancing norms.

REFERENCES

- 1) Van Bavel, J. J., Baicker, K., Boggio, P., Capraro, V., Cichocka, A., Crockett, M., ... Willer, R. (2020, March 24). Using social and behavioural science to support COVID-19 pandemic response. <https://doi.org/10.31234/osf.io/y38m9>
- 2) Lunn, P. et al. Using behavioural science to fight the coronavirus ESRI working paper no 656 (March 2020) [https://aei.pitt.edu/102644/\(2020\)](https://aei.pitt.edu/102644/(2020))
- 3) Wise, T., Zbozinek, T.D., Michelini, G., Hagan, C. C., & Mobbs, D. (2020, March 19). Changes in risk perception and protective behaviour during the first week of the COVID-19 pandemic in the United States. <https://doi.org/10.31234/osf.io/dz428>
- 4) Li, J., Yang, A., Dou, K., Wang, L., Zhang, M., & Lin, X. (2020, February 28). Chinese public's knowledge, perceived severity, and perceived controllability of the COVID-19 and their associations with emotional and behavioural reactions, social participation and precautionary behaviour: A national survey. <https://doi.org/10.31234/osf.io/5tmsh>
- 5) CDC. (2020). Coronavirus (COVID-19) <https://www.cdc.gov/coronavirus/2019-ncov/index.html>
- 6) COVID-19 Situation report 148, 2020 https://www.who.int/docs/defaultsource/coronaviruse/situation-reports/20200616-covid-19-sitrep-148-draft.pdf?sfvrsn=9b2015e9_2
- 7) India Coronavirus Worldometer, 2020, <https://www.worldometers.info/coronavirus/country/india/>.

ONLINE COURSE REPORT
ON
“DATA ANALYTICS FOR LEAN SIX SIGMA”

BY

DR BHAVNA RANA

MBA HEALTH AND HOSPITAL MANAGEMENT
SESSION 2019-21



The IIHMR University, Jaipur

1, Prabhu Dayal Marg, Sanganer Airport
Jaipur-302029, Rajasthan
Ph: 0141 3924700, Fax: 3924738
Website: www.iihmr.edu.in

COURSE NAME

“DATA ANALYTICS FOR LEAN SIX SIGMA”

COURSE DESCRIPTION

- One can get to learn different Data Analytics techniques that are typically useful within Lean Six Sigma improvement projects.
- To be able to learn to analyse and interpret the data which is gathered in a project.
- Also, to learn how to use MINITAB to analyse data gathered in such project.
- One can briefly understand what Lean Six Sigma is.
- One can learn to use Data Analytics tools and the interpretation to the outcome by using different examples from actual Lean Six Sigma projects.

One can learn techniques that can be used in a broader setting apart from Lean Six Sigma improvement projects since the data analytics tools are widely available.

OBJECTIVE OF THE COURSE

The objective of the course is to make the students learn to use the various data analytics techniques that are typically useful within Lean Six Sigma Improvement projects.

COURSE CONTENTS

This course is divided into “five-week module system” that follows different topics and their basic understanding followed by the video exercises, practice quiz and graded quiz in each week module.

The basic structure of the course follows:

WEEK 1: DATA AND LEAN SIX SIGMA

The first week introduces Lean Six Sigma and shows you where data and data analytics have their place within a DMAIC framework. It also introduces the software package Minitab. This software is used throughout the videos for data analytics.

WEEK 2: UNDERSTANDING AND VISUALIZING DATA

The second week explains how to visualise the data. It discusses visualising single variables as well as visualising two variables. One can learn to select appropriate graph. For this it is essential to first learn to distinguish between numerical and categorical data.

WEEK 3: USING PROBABILITY DISTRIBUTION

In this module on using Probability distribution, one can learn how to quantify uncertainty. Furthermore, one can also learn to answer an important business question “What percentage of products or cases meet our specifications?”

WEEK 4: (1) INTRODUCTION TO TESTING

One can learn to model their CTQ and influence factor(s) to use a decision tree to select appropriate too for data-based testing of this model. Furthermore, causality is introduced.

(2) TESTING NUMERICAL Y AND CATEGORICAL X

In this module of statistical testing, one can learn how to establish relationship between numerical Y (the CTQ) and categorical X (the influence factor).

WEEK 5: (1) TESTING NUMERICAL Y AND NUMERICAL Y

One can learn the relationship between the length of the stay and age of the patient. In this module one can learn to answer these type of questions using statistical tests to relate a numerical CTQ (the Y variable) to numerical influence factor (the X variable).

(2) TESTING: CATEGORICAL Y

Finally, one can learn to test a relationship between a Y and a X variable whenever your Y variable (the CTQ) is a categorical variable.

LEARNING METHODOLOGY

- Videos
- Lectures
- Cases
- Assignments
- Reading Material

KEY LEARNINGS OF THE COURSE

From Week 1: INTRODUCTION TO LEAN SIX SIGMA

- Process involvement is important.
- Lean Six Sigma is an important method to improve processes.
- Data is important in process improvement to improve processes by
 - Finding the biggest problem
 - Diagnosing the current process.
 - Investigating effect of improvements.
- **DMAIC**
 - Define: Select the project and establish objectives.
 - Measure: Make the problem quantifiable and measurable.
 - Analyse: Analyse the current situation and make a diagnosis.
 - Improve: Develop and implement improvement actions.
 - Control: Control the improved process performance and close the project.

From Week 2: UNDERSTANDING AND VISUALISING DATA

- **Selecting CTQs (Critical to Quality):** Property of a process, product or service that is relevant to the project objective. How to find the CTQs:
 - Which metric do I want to change?
 - Does it solve the problem?
- **CTQ is a Lean Six Sigma terminology,** in statistics this is the Y-variable, dependent variable or outcome variable.
- **Units and Operational Definitions:**
 - ✓ Units: The people, objects or phenomenon you collect data on.
 - ✓ Measurements: What a measurement means. There is never one correct operational definition, instead one 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.

The sample we collect has to be representative to the extent that the sample should be based on a mechanism that is independent of the question.

- **Organising your data:** A dataset consists of
 - Units: people, objects or phenomenon.
 - Variables: properties that are measured.
 - Units are also referred to as
 - Observational Units
 - Experimental Units
 - Cases
- **Numerical and Observational Data:**
 - Numerical Data (Quantitative)
 - ✓ Discrete: Variables that can only take integer values, such as counts (number of people, number of mistakes). Variables can be discrete because of rounding.
 - ✓ Continuous: Values not limit to integers, such as distance or temperature.
 - Categorical Data (Qualitative)
 - ✓ Ordinal Data (order): There is well defined order among categories. Eg: good, acceptable, critical, reject.
 - ✓ Nominal Data (nomen= name): Unordered categories: merely labels. Eg: XO,MM,LS,SW.
 - ✓ Binary data (bi=2): Eg: pass/fail, positive/negative, ok/not ok.
- **Distinguishing between Numerical and Categorical Data:** Numerical data contains more information than categorical data. As a rule of thumb, for a numeric CTQ, you typically need 30 observations ($n=30$). While, for categorical CTQ, you need 300 observations.
- **Visualising Data:** If Y variable of your CTQ is
 - Numerical: You can make a HISTOGRAM/BOXPLOT for it.
 - Categorical: You can make a PIECHART/BAR CHART/ PARETO CHART for it.

Histogram: used to give impression of a numerical variable.

Boxplot: It shows us the median, bulk ranges from Q1 to Q3, the whiskers (indicate the tails of the distribution), and possible outliers.

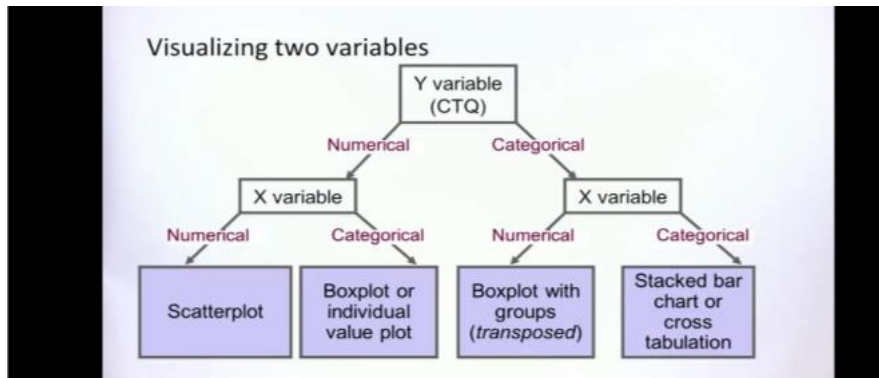
Pie-Chart: It shows us how often each category occurs.

Bar-Chart: It gives us the same information, but it is easier to spot which category occurs most often.

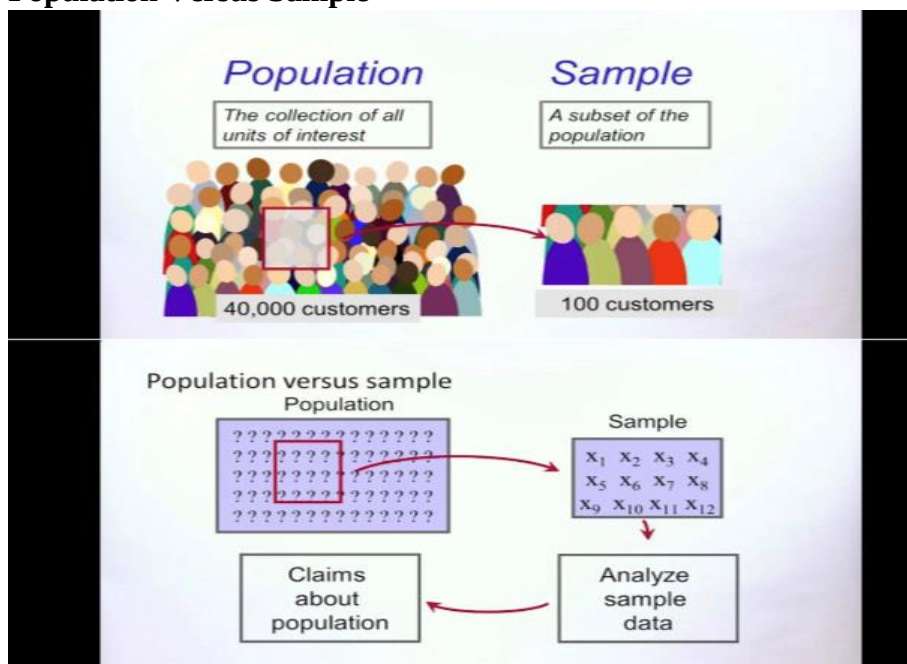
Pareto Chart: 80% of all problems can be attributed to 20% of the causes. This is Pareto's Law. A Pareto Chart can help you single out the few vital issues, from the trivial many. It can help to make a chart with another variable, like duration, to find out which problems take up most of the time.

Descriptive Statistics: While summarising the numerical data into just a few key figures, and these key figures are called Descriptive Statistics. 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.

- **Visualizing Two Variables:**



- **Population Versus Sample**



- **Estimation and Confidence Interval:**

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

From Week 3: USING PROBABILITY DISTRIBUTION

- There are hundreds of different distributions but the most frequently occurring distributions are: **NORMAL** (bell shaped), **WEIBULL** (skewed to the right) and **LOGNORMAL** (skewed distribution curve)

- **Probability Plot:** The tool used to find the best fitting distribution. We compare the AD values (Anderson Darling statistic). It measures the distance between your data and the theoretical distribution such as Normal, Weibull or Lognormal distribution.
- The lower the value, the better the fit. So, we select the distribution within the lowest AD value.
- **Empirical CDF:** It is used to determine how many percent of the cases meet your service level agreement, or your specification.

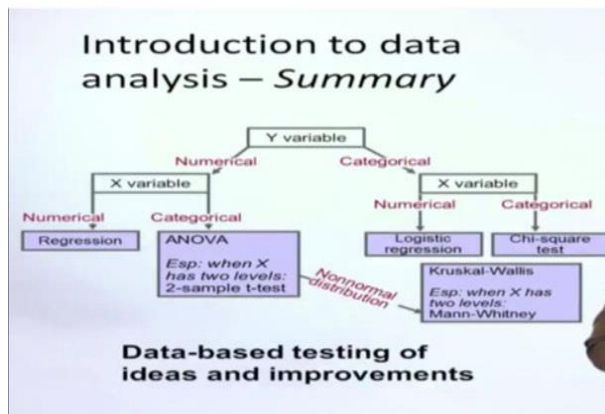
From Week 4 & 5: Introduction to Testing

Testing: Numerical Y and categorical X.

Testing: Numerical Y and Numerical X

Testing: Categorical Y

- Data-based testing of ideas and improvements. This is one of principles of Lean Six Sigma. This principle is sometimes referred to as evidence-based improvement actions.
- To analyse if two variables are related, you must determine which variable is your CTQ, or your Y.
- Once you have determined what type of data you have, the tree diagram can be very useful tool to determine which type of analysis method is suitable for your data.



- **Hypothesis Testing:** Testing means that we claim something. Ans we use data to test the claim. This claim is called a hypothesis. You always begin with a null hypothesis, which is the hypothesis that there is an effect between variables. Next you must collect evidence, which will be mostly to calculate your p-value. This p-value allows you to decide whether or not to reject your null hypothesis. We calculate p-value or probability value, if p-value is larger than 0.05 then we neglect the null hypothesis, and if p-value is larger than 0.05, then we accept null hypothesis.
- **Causality:** Its 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.
Possible errors when making casual assumption:
 - Wrong direction.
 - Time problem
 - Third variable problem
 - Underlying variable problem.
- **Introduction to Anova:** It stands for analysis of variance. It is an appropriate statistical technique when the Y variable is numerical, and the X variable is categorical. And it tests whether these groups have equal means or not.
- **Kruskal-Wallis Test:** It 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.

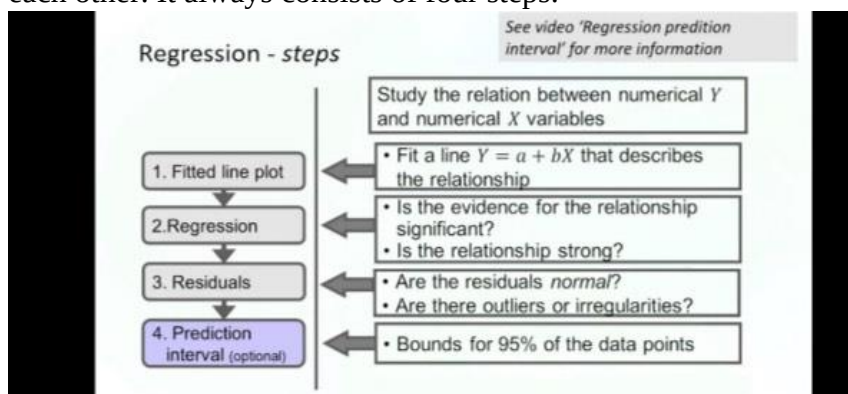
- **Two Sample T-test:** It is a simple test to compare the means of two different groups. The p-value can be used to calculate whether the difference of the two means are coincidental or truly significant. The difference between one-way ANOVA and two sample t-test is that one way ANOVA is used to compare means of more than 2 groups at a single time.
- **Test for equality of variances:** Increasing variability always degrade the performance of a production system. This is a law of variability.
- **Correlation:** It 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.

Correlation - Overview

Rule of thumb:

0 to 0.4	No or weak positive correlation
0.4 to 0.8	Moderate positive correlation
0.8 to 1	Strong positive correlation
0 to -0.4	No or weak negative correlation
-0.4 to -0.8	Moderate negative correlation
-0.8 to -1	Strong negative correlation

- **Introduction to Regression:** It is a statistical method to test if two numerical variables are related to each other. It always consists of four steps:



Next step after regression analysis is residual analysis. We must check if the residuals are normally distributed and if the residuals do not show any outliers or other regularities.

- **Quadratic Regression:** It fits a curved line. After the main analysis, you have to perform a residual analysis.
- **Chi-Square Analysis:** It is a suitable method if you have a categorical Y and a categorical X variable. The first step is to make cross tabulation. The second step is to look at the p-value and check if your results are significant.
- **Logistic Regression Analysis:**

