

**IMPACT OF COVID 19 LOCKDOWN ON
HEALTH RISK BEHAVIOUR OF INDIVIDUALS**

A Report by

Preetha P

Guided by

Dr Arindam Das

MBA Hospital and Health Management

2019-2021

ACKNOWLEDGEMENTS

The journey to become a wise and responsible professional does not confine to classroom learnings. The experience that I gained while working with this study and making of the report, acquainted me with the practical aspect of Research Methodology & implementation of the theory in the real life. Any compliments require the effort of the respondents to the survey for their valuable time in filling up the form. My sincere thanks to all for taking me ahead in this journey.

I am appreciative to my guide Dr Arindam Das, Associate Professor, Institute of Health Management Research, Jaipur for the consistent help identified with my questions during study, for his understanding, inspiration, and enormous information. His direction helped me in all the time recorded as a hard copy of this report.

I would also like to thank my family and colleagues for their constant support and for keeping me motivated throughout the journey.

TABLE OF CONTENTS

S.No	Content	Page Number
1	Introduction	9
2	Material and Methods	10
3	Results	11
3.1	Demographic responses of respondents	11
3.2	Work status and number of persons in household	11
3.3	Pattern of health risk behaviour and frequency in lockdown	12
3.4	Association between characteristics & smoking pattern in the lockdown	14
3.5	Association between characteristics & alcohol consumption pattern in the lockdown	17
3.6	Reasons for smoking less in the lockdown	19
3.7	Association between characteristics & smoking less in the lockdown	20
3.8	Reasons for smoking more in the lockdown	21
3.9	Association between characteristics & smoking more in the lockdown	21
3.10	Getting cigarettes in the lockdown	22
3.11	Reasons for consuming alcohol less during lockdown	22
3.12	Association between characteristics & consuming alcohol less during lockdown	23
3.13	Reasons for consuming alcohol more during lockdown	24
3.14	Association between characteristics & consuming alcohol more during lockdown	24
3.15	Getting alcohol in the lockdown	25
3.16	Attitude of respondents after lockdown	26
3.17	Association between characteristics & changing attitude after lockdown	26
4	Discussion	27
5	Conclusion	29
6	References	29
7	Annexure	30

LIST OF TABLES

S.no	Content	Page Number
1.	Demographic responses of respondents	11
2.	Work status and number of persons in household	12
3.	Association between characteristics & Smoking pattern	14
4.	Association between characteristics & Drinking pattern	17
5.	Reasons for smoking less in the lockdown	19
6.	Association between characteristics & Reasons on smoking less in the lockdown	20
7.	Reasons for smoking more in the lockdown	21
8.	Association between characteristics & Reasons on smoking more in the lockdown	21
9.	Getting cigarettes in the lockdown	22
10.	Reasons for consuming alcohol less during lockdown	23
11.	Association between characteristics & Reasons on consuming alcohol less during lockdown	23
12.	Reasons for consuming alcohol more during lockdown	24
13.	Association between characteristics & Reasons on consuming alcohol more during lockdown	24
14.	Getting alcohol in the lockdown	26
15.	Attitude of respondents after lockdown	26
16.	Association between characteristics & Change of Behaviour after lockdown	27

LIST OF FIGURES

S.no	Content	Page Number
1	Smoking pattern V/s Frequency	12
2	Drinking pattern V/s Frequency	13
3	Smoking pattern during lockdown	13
4	Drinking pattern during lockdown	13

LIST OF ABBREVIATIONS

COVID	Corona Virus Disease
SARS	Severe Acute Respiratory Syndrome
CoV	Corona Virus
N	Number of Sample

ABSTRACT

BACKGROUND: COVID 19 has been declared as a public health emergency, that is affecting the globe. Specialists worldwide have reacted by actualising travel limitations, lockdowns, work environment risk controls, and office terminations.⁽¹⁾ Securing amidst pandemic likewise remembered boycotts for cigarettes and liquor deals. A few nations over the world have implemented a severe prohibition on tobacco and liquor deals during the lockdown in a manner influencing the utilisation patterns of alcohol and tobacco consumption of the purchasers.⁽²⁾ Lushes and smokers are thinking that it's difficult to take their day by day portion of nicotine as quarantine has secured them with their family, while some are wrestling with withdrawals; others despite everything are utilising the opportunity to kick the butt.⁽⁴⁾

OBJECTIVE: Broad Objective- To study impact of COVID -19 lockdown on health risk behaviours (alcohol consumption and smoking) of individuals. **Specific Objective-** To study if alcohol consumption and smoking has been more than usual, same as usual or less than usual since the lockdown and analyse the trend based on gender, age group, place of residence, work, and household. To study the reasons behind change in trends of drinking alcohol and smoking since the lockdown and analysis based on gender, age group, place of residence, work, and household. To know if the individuals are expecting to have a change in their consumption habits post lockdown and analyse the trend based on gender, age group and place of residence, work, and household

RESULTS: A total of 206 responses were received among which 119 were complete responses for smoking (response rate 58%) and 90 were complete responses for alcohol consumption (response rate 44%) that were taken for analysis. 50.42% respondents who smoke perceived their smoking pattern to be more than usual, 41.18% perceived it as less than usual and 8.40% perceived their smoking pattern to be same as usual during lockdown. On the other hand, 57.78% of the respondents who consume alcohol perceived their drinking pattern to be less than usual, 15.56% perceived it to be more than usual whereas 26.67% of the respondents perceived their drinking pattern to be same as usual during the lockdown. A statistical difference was found in the perceived smoking pattern of respondents with respect to gender ($p = 0.031$), age- group ($p = 0.004$), residence ($p = 0.000$), work- status ($p = 0.000$) and household ($p = 3.86 \times 10^{-6}$). A statistical difference in the alcohol consumption pattern during lockdown was found only with respect to place of residence ($p = 0.000$) and number of persons in the household ($p = 0.02$). It was found that 44.88% respondents were willing to reduce their health risk behaviour after lockdown as they faced positive outcomes on their physical and mental health whereas, 55.12% respondents did not consider changing their health risk behaviour.

CONCLUSION: The study helped in knowing the patterns of consumption design amidst lock down and in tossing experiences regarding the psychological conduct and approach of people towards such health risk hazard practices. Individuals who are self-driven thinking lockdown to be a best time to quit alcohol consumption and smoking should be further motivated on quitting these "habit driven" behaviours post lockdown and reduce the non-communicable disease burden.

1. INTRODUCTION

The COVID-19 pandemic, is viewed as the pandemic of coronavirus malady 2019 (COVID 19), brought about by severe acute respiratory syndrome coronavirus 2 (SARS CoV 2).⁽¹⁾ The flare-up was first recognised in Wuhan, China, in December 2019.⁽¹⁾ The World Health Organisation announced the flare-up a Public Health Emergency of International Concern on 30 January, and a pandemic on 11 March.⁽¹⁾ The infection is fundamentally spread between individuals during close contact, regularly through little droplets created by hacking, sniffing, and talking.⁽¹⁾ The droplets as a rule tumble to the ground or onto surfaces as opposed to going through air over significant distances.⁽¹⁾ Less regularly, individuals may get tainted by contacting a sullied surface and afterwards contacting their face.⁽¹⁾ It is generally infectious during the initial three days after the beginning of side effects, albeit spread is conceivable before manifestations show up, and from individuals who do not show symptoms.⁽¹⁾ Basic symptoms incorporate fever, cough, exhaustion, brevity of breath, and loss of smell.⁽¹⁾ Intricacies may be pneumonia and intense respiratory trouble.⁽¹⁾ The time from presentation to beginning of side effects is regularly around five days yet may go from two to fourteen days.⁽¹⁾ There is no known immunisation or explicit antiviral treatment.⁽¹⁾

Primary treatment is supportive therapy.⁽¹⁾ Suggested preventive measures incorporate hand washing, covering one's mouth when coughing, keeping up social distancing from others, wearing a face veil in open settings, and observing and self-confinement for individuals who presume they are tainted.⁽¹⁾ Specialists worldwide have reacted by actualising travel limitations, lockdowns, work environment risk controls, and office terminations.⁽¹⁾ Numerous spots have additionally attempted to expand testing limits and follow contacts of contaminated people.⁽¹⁾ The pandemic has caused worldwide social and financial disturbance, including the biggest worldwide downturn since the Great Depression.⁽⁴⁾ It has prompted the deferment or dropping of donning, strict, political, and social occasions, across the board gracefully deficiencies exacerbated by alarm purchasing, and diminished outflows of toxins and ozone harming substances.⁽²⁾ Schools, colleges, and universities are right now shut either on an across the nation or neighbourhood premise in 177 nations, influencing around 98.6 level of the world's understudy populace.⁽¹⁾

Securing amidst pandemic likewise remembered boycotts for cigarettes and liquor deals. A few nations over the world have implemented a severe prohibition on tobacco and liquor deals during the lock-down in a manner influencing the utilisation patterns of alcohol and tobacco consumption of the purchasers.⁽²⁾ The Covid-19 activated lock down has claimed a substantial cost for all the areas of the economy and the 'transgression segment' is no exemption.⁽²⁾ An expansion in cost of tobacco and liquor does not influence the interest for it.⁽²⁾ People who enjoy any of the wellbeing hazard practices, be it liquor consumption or cigarette smoking do as such for nervousness help and stress busting.⁽²⁾ Remembering this, numerous nations like Switzerland, France, Italy and Spain have permitted the offer of cigarettes and liquor during the lockdown and let little shops stay open so individuals don't need to experience withdrawal indications and adapt to extra nervousness while previously adapting to the pressure of the pandemic.⁽²⁾ Be that as it may, no such special cases were made in India, where many have needed to manage the inaccessibility of cigarettes and liquor.⁽³⁾ Lushes and smokers are thinking that it's difficult to take their day by day portion of nicotine as quarantine has secured them with their family, while some are wrestling with withdrawals; others despite everything are utilising the opportunity to kick the butt.⁽⁴⁾ Knowing the patterns of consumption design amidst lock down gets enthusiasm for tossing experiences in regards

to the psychological conduct and approach of people towards such health risk hazard practices and if a pandemic with lock-down can acquire constructive results in people enjoying cigarette smoking and liquor consumption in this way decreasing the non-communicable disease burden.

2. MATERIALS AND METHODS

Prior to the implementation of this study, the questionnaire structure, validation, and strategies to enrol individuals who practice health risk behaviours for maximum participation across the globe was discussed. A research proposal was drafted, and the study protocol was followed as mentioned in the proposal, approved by Dr Arindam Das – Associate Professor IIHMR University, Jaipur. The mode of data collection was through the link : <https://docs.google.com/forms/d/16L28P71gcoOilTJW3jPUprmFzJpWeysWVRRqYb6tI/edit> . Upon clicking the form, the first page described the objective of the research which was to study if the Health risk behaviours like alcohol consumption and smoking have been more than usual, same as usual or less than usual in the lockdown and the reasons behind change in trends with an analysis based on gender, age group, place of residence, work and household. The consent page of the questionnaire promised on keeping the confidentiality of the participants and that their participation in the study was voluntary. The inclusion criteria for the participants was to have access to internet connection to obtain the questionnaire. Participants having any one of the Health risk behaviours between the age of 18 years - 67 years were included in the study. The exclusion criteria included participants who did not give complete responses of the parts of questionnaire. The duration of the study was from 13th May - 13th June. The questionnaire was purposefully forwarded to individuals who practiced any one of the Health risk behaviours. The questionnaire was broadcasted to friends and relatives staying in India and Abroad who smoke and consume alcohol. It was also shared on social media platforms like LinkedIn and Facebook. To meet the required sample size of 185 which was estimated by (Z^2pq/d^2) and maximum participation, one of the owners of a Resto bar was requested to provide with mail IDs of customers who visited the bar in the month of February and the link for the questionnaire was forwarded to them.

The questionnaire focused on aspects like respondent's profile (gender, age group, place of residence), respondent's work status and number of persons living in the household, respondent's health risk behaviour and respondent's attitude after lockdown. The questionnaire consists of 20 questions. Question 1-3 focuses on the profile of the respondent. Questions 4-6 focuses on the work and household of the respondent. Question 7 focuses on the health risk behaviour of the respondent. Respondents who only smoke answer to Part I of the questionnaire (Question 8-12). Respondents who only consume alcohol answer Part II of the questionnaire (Question 13-18). Respondents who indulge in both the health risk behaviour answer to both the parts of the questionnaire. Question 19-20 focuses on the attitude of the respondents after the lockdown.

The data analysis included descriptive statistics of the percentages of individuals under each variable and a Chi-square analysis of the consumption pattern of smoking and drinking alcohol with respect to Gender, Age- group, Residence, Work status and Number of persons in the household was done. Analysis on the perceptive reasons chosen for a change in pattern (less than usual or more than usual in the lockdown) was also projected. The analysis was done using Ms -Excel, R and Minitab Statistical Software.

3. RESULTS

A total of 206 responses were recorded among which 119 were complete responses for smoking (response rate 58%) and 90 were complete responses for alcohol consumption (response rate 44%) that were taken for analysis.

3.1 Demographic characteristics of the respondents

The total number of responses were divided based on respondent's health risk behaviour practiced and the analysis was done. The demographic characteristics has been presented in Table 1. Of the total respondents, 119 had a habit of smoking in which 79% were males and 21% were females, similarly, 90 respondents had a habit of consuming alcohol in which 77% were males and 23% were females. 69% of the respondents who smoke were under the age group of 18-27 years and 21% of the respondents fell under the age group of 28-37 years. 18-27 years respondents showed a greater percentage of alcohol consumption (60%) whereas 33% of the respondents who drink alcohol were 28-37-year-old. Since this study had a global approach, the smoking and drinking pattern seen in India were 67%, 49% respectively. The smoking and drinking pattern Abroad were seen to be 33% and 51% respectively.

Table 1: Demographic characteristics of respondents

Smoking	Characteristics	N	N%
Gender	Female	25	21
	Male	94	79
Age Group	18-27 years	82	69
	28-37 years	25	21
	38-47 years	6	5
	48-57 years	6	5
	58-67 years	0	0
Residence	India	80	67
	Abroad	39	33
Alcohol	Characteristics	N	N%
Gender	Female	21	23
	Male	69	77
Age-group	18-27 years	54	60
	28-37 years	30	33.33
	38-47 years	5	5.56
	48-57 years	1	1.11
	58-67 years	0	0
Residence	India	46	49
	Abroad	44	51

3.2 Work status and Number of persons in household during lockdown of the respondents

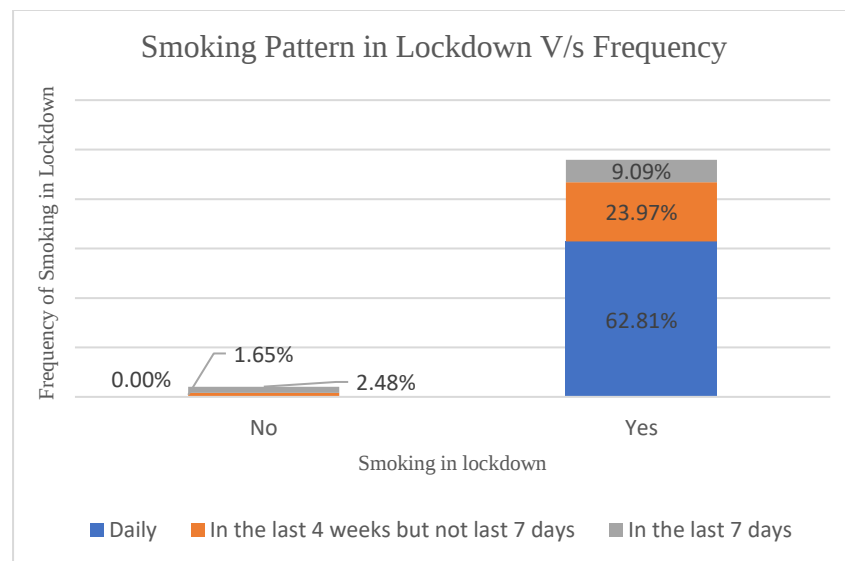
It was observed in Table 2 that during lockdown, 61% of the respondents who smoke and 78.89% respondents who drink have been working even during the lockdown. Most of the respondents live in a household with 2-4 persons, 58.8% of the smoke and 50% of them consume alcohol.

Table 2: Work status and household of respondents

Smoking	Characteristics	N	N%
Work status	Yes	73	61
	No	46	39
Household	Living alone	40	33.6
	2-4 persons	70	58.8
	5-6 persons	8	6.7
	more than 7 persons	1	0.8
Alcohol	Characteristics	N	N%
Work status	Yes	71	78.89
	No	19	21.11
Household	Living alone	36	40
	2-4 persons	45	50
	5-6 persons	8	8.89
	more than 7 persons	1	1.11

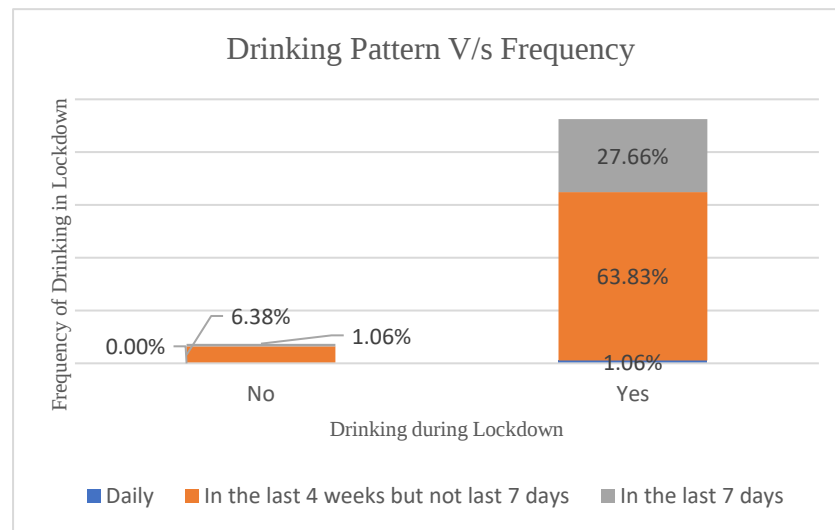
3.3 Pattern of health risk behaviour and frequency in the lockdown

Figure 1: Smoking pattern in lockdown V/s Frequency



As observed in Figure 1 , more than two thirds (95.87 %) of the respondents continued smoking during lockdown, out of which 62.81 % smoked daily, 23.97% smoked in the 4 weeks but not anytime recently and 9.09% of the respondents smoked in the last 7 days.

Figure 2: Drinking pattern in lockdown V/s Frequency



Looking at the alcohol consumption pattern in Figure 2 , 92.55% of the respondents continued drinking alcohol during lockdown, out of which, only 1.06% of the respondents drank daily, 63.83% drank in the last 4 weeks but not recently and 27.66% drank in the last 7 days .

Figure 3: Smoking pattern during lockdown

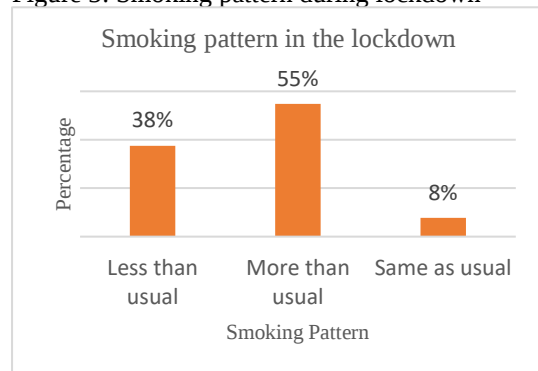
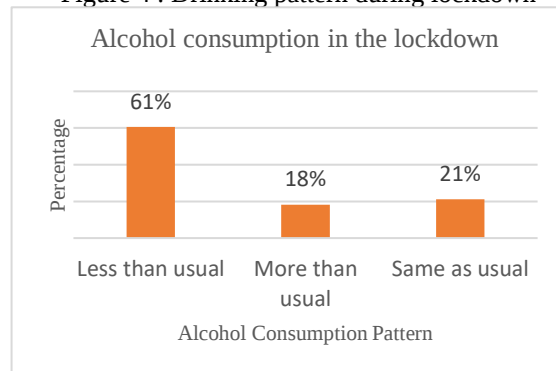


Figure 4 : Drinking pattern during lockdown



The health risk behaviour pattern in the lockdown had been presented in Figure 3 and Figure 4. As observed, 55% respondents who smoke perceived their smoking pattern to be more than usual, 38% perceived it as less than usual and 8% perceived their smoking pattern to be same as usual during lockdown. On the other hand, less than two thirds (61%) of the respondents who consume alcohol perceived their drinking pattern to be less than usual, 18% perceived it to be more than usual whereas 21% of the respondents perceived their drinking pattern to be same as usual during the lockdown.

3.4 Association between characteristics of the respondents with Smoking pattern

Table 3: Association between characteristics of the respondents with Smoking pattern

		<i>Less than usual</i>	<i>More than usual</i>	<i>Same as usual</i>	<i>p-value</i>
Gender					
	Female	7	18	0	
		28	72	0	
		14.29	30	0	
	Male	42	42	10	0.031*
		44.68	44.68	10.64	
		85.71	70	100	
	All	49	60	10	
		41.18	50.42	8.4	
		100	100	100	
Age group					
	18-27 years	39	37	6	
		47.56	45.12	7.32	
		79.59	61.67	60	
	28-37 years	6	18	1	
		24	72	4	
		12.24	30	10	
	38- 47 years	1	5	0	0.0004*
		16.67	83.33	0	
		2.04	8.33	0	
	48-57 years	3	0	3	
		50	0	50	

		6.12	0	30	
	All	49	60	10	
		41.18	50.42	8.4	
		100	100	100	
Residence					
	Abroad	3	32	4	
		7.69	82.05	10.26	
		6.12	53.33	40	
					0.000*
	India	46	28	6	
		57.5	35	7.5	
		93.88	46.67	60	
	All	49	60	10	
		41.18	50.42		
		100	100		
Work status	No	32	9	5	
		69.57	19.57	10.87	
		65.31	15	50	
	Yes	17	51	5	0.000*
		23.29	69.86	6.85	
		34.69	85	50	
	All	49	60	10	
		41.18	50.42	8.4	
		100	100	100	
	Living alone	3	34	3	

Household		7.5	85	7.5	
		6.12	56.67	30	
	2-4 persons	38	25	7	
		54.29	35.71	10	
		77.55	41.67	70	
	5-6 persons	7	1	0	3.8*10 ^{-6*}
		87.5	12.5	0	
		14.29	1.67	0	
	more than 7 persons	1	0	0	
		100	0	0	
		2.04	0	0	
	All	49	60	10	
		41.18	50.42	8.4	
		100	100	100	
	Cell Contents				
	Count				
	% of Row				
	% of Column				

*p<0.05

As observed in Figure 3, the smoking pattern was perceived to be more than usual by the respondents. A chi square test was carried out to find out the significance of smoking pattern with respondent's profile. Results of the same are being represented in Table 3. It was emerged from the analysis that all the variables like gender, age- group, place of residence, working status and number of persons in the household influenced the smoking pattern in the lockdown significantly.

3.5 Association between characteristics of the respondents with alcohol consuming pattern

Table 4: Association between characteristics of the respondents with alcohol consuming pattern

		<i>less than usual</i>	<i>more than usual</i>	<i>same than usual</i>	<i>p-value</i>
Gender	Female	12	1	8	0.18
		57.14	4.76	38.1	
		23.08	7.14	33.33	
	Male	40	13	16	
		57.97	18.84	23.19	
		76.92	92.86	66.67	
	All	52	14	24	
		57.78	15.56	26.67	
		100	100	100	
Age group	18-27 years	29	11	14	0.22
		53.7	20.37	25.93	
		55.77	78.57	58.33	
	28-37 years	18	3	9	
		60	10	30	
		34.62	21.43	37.5	
	38- 47 years	5	0	0	
		100	0	0	
		9.62	0	0	
	48- 57 years	0	0	1	
		0	0	100	
		0	0	4.17	
	All	52	14	24	

		57.78	15.56	26.67	
		100	100	100	
Residence	Abroad	22	2	20	
		50	4.55	45.45	
		42.31	14.29	83.33	
					0*
	India	30	12	4	
		65.22	26.09	8.7	
		57.69	85.71	16.67	
	All	52	14	24	
		57.78	15.56	26.67	
		100	100	100	
Work status	No	11	2	6	
		57.89	10.53	31.58	
		21.15	14.29	25	
					0.737
	Yes	41	12	18	
		57.75	16.9	25.35	
		78.85	85.71	75	
	All	52	14	24	
		57.78	15.56	26.67	
		100	100	100	
Household	Living alone	25	0	11	
		69.44	0	30.56	
		48.08	0	45.83	
	2-4 persons	20	13	12	
		44.44	28.89	26.67	
		38.46	92.86	50	
					0.02*
	5-6 persons	6	1	1	

	75	12.5	12.5	
	11.54	7.14	4.17	
more than 7 persons	1	0	0	
	100	0	0	
	1.92	0	0	
All	52	14	24	
	57.78	15.56	26.67	
	100	100	100	
Cell Contents				
Count				
% of Row				
% of Column				

*p<0.05

As observed in Figure 4, the alcohol consumption pattern among respondents during lockdown was perceived to be less than usual. A chi square test was carried out to find out the significance of alcohol consumption pattern with respondent's profile. Results of the same are being represented in Table 4. It was emerged from the analysis that variables like place of residence and number of persons in the household influenced the alcohol consumption pattern in the lockdown significantly.

3.6 Reasons for smoking less than usual during the lockdown

A question was asked to record the reasons they felt that led to reduce smoking during lockdown. As observed in Table 5, 85.71% of the respondents reported cigarettes were not available during lockdown, as a result they smoked less. Another 46.94% of the respondents mentioned less or no social life lead to reduced smoking pattern during lockdown. More than one-third (38.78%) of the respondents felt the lockdown was a good period to reduce smoking and hence themselves reduced their smoking consumption.

Table5: Reasons for smoking less during lockdown

	N%
Reasons	
Financial constraints	12.24%
No cigarettes available	85.71%
No social life	46.94%
Chances of getting COVID 19	8.16%

Living in Smoke free household	32.65%
Suggestions by others to reduce	8.16%
Increased workload	0%
Self-reduction	38.78%

3.7 Association between characteristics and less smoking

A chi square test was carried out to find out the significant reasons with less smoking during lockdown. Results of the same are being represented in Table 6. It was emerged from the analysis that financial constraints among individuals living with 5-6 members in their household reduced their smoking amongst significantly. No or less social life was observed to have significantly reduced the smoking pattern amongst males. 38-47 years age group felt that smoking would increase their chances of getting COVID 19, which significantly reduced their smoking pattern.

Table 6: Association between characteristics and less smoking

		FINANCIAL CONSTRAINT S%	NO CIGARETTE S%	NO SOCIAL LIFE %	CHANCES OF GETTING COVID %	SMOKE FREE HOUSEHOLD %	SUGGESTIONS BY OTHERS TO REDUCE%	INCREASE WORK LOAD %	SELF REDUC TION%
Gender	Female	0	100	14.29	0	28.57	14.29	NA	28.57
	Male	14.29	83.33	52.38	9.52	33.33	7.14	NA	40.48
	p-value	0.65	0.24	0.01*	0.91	1	1	NA	0.85
Age group	18-27 years	12.8	87.2	41.03	7.69	38.5	10.3	NA	35.9
	28-37 years	16.7	66.7	50	0	16.7	0	NA	33.3
	38-47 years	0	100	100	100	0	0	NA	0
	48-57 years	0	100	100	0	0	0	NA	100
	58-67 years	0	0	0	0	0	0	NA	0
	p-value	0.87	0.47	0.51	0.07*	0.35	0.77	NA	0.13
Residence	Abroad	0	100	100	0	0	0	NA	0
	India	13	84.78	43.48	8.7	34.78	8.7	NA	41.3
	p-value	1	1	0.16	1	0.54	1	NA	0.4
Work status	working	11.7	88.24	41.18	11.76	35.29	11.76	NA	29.41
	not working	12.5	84.38	50	6.25	31.25	6.25	NA	43.75
	p-value	0.94	1	0.35	0.90	1	0.90	NA	0.50
Household	Living alone	33.33	66.67	33.33	33.33	0	0	NA	0
	2-4 persons	2.63	86.84	52.63	5.26	34.2	7.89	NA	47.4
	5-6 persons	57.14	85.71	14.29	14.29	28.6	14.29	NA	14.3
	more than 7 persons	0	100	4.35	0	100	0	NA	0
	p-value	0.004*	0.73	0.48	0.33	0.30	0.87	NA	0.13

*p<0.05 NA Not applicable

3.8 Reasons for smoking more than usual during the lockdown

A question was asked to record the reasons with more smoking during lockdown. Results of which are presented in Table 7. As observed, 75.86% respondents reported boredom driving them to smoke more than usual in the lockdown. Another 72.41% took on to smoking more than usual to relax themselves in the COVID 19 lockdown. Half (50%) of the respondents have been feeling stressed out and anxious eventually taking on to more cigarettes. Less than one third of the respondents (24.14%) have been socialising with people at home driving them in smoking more than usual during lockdown.

Table 7: Reasons for more smoking during lockdown

<i>Reasons</i>	<i>N%</i>
<i>feeling bored in the lockdown</i>	75.86%
<i>way to relax</i>	72.41%
<i>feeling stressed out</i>	50%
<i>Smoking more with people at home</i>	24.14%
<i>Online occasion</i>	0%
<i>while meeting friends outside</i>	8.62%

3.9 Association between characteristics and more smoking

A chi square test was carried out to find out the significant reasons with more smoking during lockdown. Results of the same are being represented in Table 8. It was emerged from the analysis that boredom was significantly a reason to smoke more amongst the middle-aged group and in India. Smoking was taken as a mode of relaxation amongst individuals in the early 20s and 30s who have been working in the lockdown. Those living alone also stated the above reason significantly associated with increased smoking pattern. More than two-thirds of the females feel stressed out in the lockdown making them smoke more amidst the pandemic. Stress was also seen as a significant reason for the working population and more than half of the individuals living alone in the lockdown. It emerged significantly that more than one third of the males in the early 20s living with less than 4 members in their household have been smoking more with people at their homes taking on to an increased smoking pattern. The above reason has also been found to be significantly associated with individuals not working in the lockdown. Smoking more while meeting friends outside was found as a significant reason to smoke more among individuals in their early 30s.

Table 8: Association between characteristics and more smoking

		FEELING BORED IN THE LOCKDOWN %	WAY TO RELAX%	FEELING STRESSED OUT%	SMOKING WITH PEOPLE AT HOME%	ONLINE OCCASION%	WHILE MEETING FRIENDS OUTSIDE %
GENDER	Female	72.22	83.33	83.33	0	NA	0
	Male	77.5	67.5	35	35	NA	12.5
	p-value	0.66	0.21	0.001*	0.004*	NA	0.2
	18-27 years	61.1	69.4	44.44	38.9	NA	0

	28-37 years	100	100	70.59	0	NA	23.53
AGE GROUP	38-47 years	100	0	20	0	NA	20
	48-57 years	0	0	0	0	NA	0
	58-67 years	0	0	0	0	NA	0
	p-value	0.004*	0.00*	0.07	0.003*	NA	0.01*
RESIDENCE	Abroad	100	68.75	62.5	18.75	NA	12.5
	India	46.12	76.92	34.62	30.77	NA	3.85
	p-value	0.00*	0.48	0.03*	0.28	NA	0.24
WORK STATUS	working	74	82	44	16	NA	0
	not working	87.5	12.5	87.5	75	NA	10
	p-value	0.4	0.00*	0.02*	0.000*	NA	0.79
	Living alone	85.2	94.12	58.82	0	NA	11.76
	2-4 persons	65.22	39.13	39.13	60.9	NA	4.35
HOUSEHOLD	5-6 persons	0	0	0	0	NA	0
	more than 7 persons	0	0	0	0	NA	0
	p-value	0.91	0.0000255*	0.2*	0.0000007*	NA	0.59

*p<0.05 NA Not applicable

3.10 Responses on buying cigarettes during lockdown

A question was asked to record the reasons on how people had been getting hold of cigarettes even after the imposed lockdown. The results have been presented in Table 9. As observed, more than two-thirds (80%) of the population have been still buying tobacco to meet their daily nicotine requirements from local stores and supermarkets. There were less than one third of population (31.78%) who had stocked up cigarettes before the imposition of the lockdown and another one third have been offered by friends or relatives.

Table 9: Buying cigarettes in the lockdown

	N%
<i>Bought enough supply</i>	31.78
<i>Buying on demand</i>	80
<i>Offered from elsewhere</i>	31.78

3.11 Reasons for consuming alcohol less than usual during lockdown

Another question was asked to record the reasons on why individuals felt their alcohol consumption was less than usual in the lockdown. The results have been presented in Table 10. As observed, more than two thirds (78.8%) of the population felt less or no social life as one of the main reasons for decreased consumption. Another, two thirds (65.3%) found difficulty in getting hold of alcohol with the on- going restrictions and shops being closed.

Less than one third of the population lived in a smoke- free household and did not like drinking around family members. Individuals who were self- driven to reduce their alcohol consumption considering lockdown to be the best time to reduce their alcohol consumption accounted to only 13.4 %.

Table 10: Reasons for consuming less alcohol during lockdown

<i>Reasons for drinking less in the lockdown</i>	<i>N%</i>
No social life	78.85
Financial constraints	9.62
Difficult to get alcohol	65.38
Do not like drinking at home	15.38
Someone suggested on reducing	5.77
Increase workload	3.85
Self-reduction	13.46

3.12 Association between characteristics and less alcohol consumption

A chi-square test was carried out to find the significant reason for drinking less alcohol in the lockdown. The results have been presented in Table 11. It was emerged from the analysis that no or less social life was a significant reason to drink less alcohol amongst individuals Abroad. Financial constraints in the lockdown was seen to be a significant reason for drinking alcohol less amongst individuals living in India. Females found difficulty in getting hold of alcohol in the lockdown as a significant reason to drink less, this was also the case with more than two thirds of the individuals of age group 18-27 years and those who have been working in the lockdown. Indians who live with more than 2 persons in their household don't find it comfortable drinking around family members and this has been seen as a significant reason in reduced alcohol consumption in the lockdown.

Table 11: Association between characteristics and less alcohol consumption

		<i>NO SOCIAL LIFE %</i>	<i>FINANCIAL CONSTRAINTS %</i>	<i>DIFFICULT TO GET ALCOHOL%</i>	<i>NO DRINKING AT HOME %</i>	<i>SUGGESTIONS BY OTHERS TO REDUCE %</i>	<i>INCREASE WORK LOAD%</i>	<i>SELF REDUCTION%</i>
Gender	Female	91.67	8.33	41.67	16.67	8.33	0	16.67
	Male	26.83	10	72.5	15	5	5	12.5
	p-value	0.21	0.8	0.04*	0.88	1	1	0.7
Age group	18-27 years	68.97	10.34	65.52	24.14	10.34	6.94	16.94
	28-37 years	88.99	11.11	55.52	5.56	0	0	0
	38-47 years	100	0	100	0	0	0	0
	48-57 years	0	0	0	0	0	0	0
	58-67 years	0	0	0	0	0	0	0
	p-value	0.12	0.74	0.1	0.13	0.2	0.43	0.54

Residence	Abroad	95.45	0	63.64	0	0	0	48.64
	India	66.67	16.67	66.64	26.67	10	6.67	12.02
	p-value	0.01*	0.04*	0.8	0.008*	0.12	0.61	0.22
Work status	working	82.93	7.32	70.73	12.2	7.32	0	7.32
	not working	63.64	18.18	45.45	27.27	0	4.88	4
	p-value	0.16	0.27	0.18	0.21	0.84	1	0.31
Household	Living alone	88	8	64	0	0	0	33.33
	2-4 persons	70	5	70	35	10	5	0
	5-6 persons	83	33.33	50	16.7	16.67	16.67	0
	more than 7 persons	0	0	100	0	0	0	0
	p-value	0.119	0.20	0.71	0.013*	0.31	0.28	0.18

*p<0.05 NA Not Applicable

3.13 Reasons for consuming alcohol more than usual during lockdown

A question was asked to know the reasons for drinking more alcohol in the lockdown. The results have been presented in Table 12. As observed, more than two thirds (75%) of the population took on to drinking more alcohol than usual in the lockdown as they have been stressed out in this pandemic and the imposed lockdown has made them more social with people at home creating an environment of their own with members of their household driving them to drink more alcohol than usual in the pandemic.

Table 12: Reasons for consuming more alcohol

Reasons for drinking more in the lockdown	N%
decrease chances of getting COVID 19	0
Help in relaxing	16.67
Feeling bored	16.67
Feeling stressed out	75
Being more social at home with family	75
Need not get up for work	0
Online occasion	0

3.14 Association between characteristics and more alcohol consumption

A chi-square test was performed to find the significant reason for consuming alcohol more than usual during the lockdown. The results are presented in Table 13 and it was emerged from the analysis that the reasons mentioned by individuals in the table below were not significant enough for a conclusion.

Table 13: Association between characteristics and more alcohol consumption

		<i>Decrease chances of getting COVID 19 %</i>	<i>Help in relaxing %</i>	<i>Feeling bored %</i>	<i>feeling stressed out %</i>	<i>Being more social at home %</i>	<i>Needn't get up for work %</i>	<i>Online occasion%</i>
Gender	Female	NA	0	0	100	0	NA	NA
	Male	NA	18.18	18.18	72.73	81.82	NA	NA
	p-value	NA	1	1	1	0.54	NA	NA
Age group	18-27 years	NA	11.11	11.11	88.89	88.89	NA	NA
	28-37 years	NA	33.33	33.33	33.33	33.33	NA	NA
	38-47 years	NA	0	0	0	0	NA	NA
	48-57 years	NA	0	0	0	0	NA	NA
	58-67 years	NA	0	0	0	0	NA	NA
	p-value	NA	1	1	0.2	0.24	NA	NA
Residence	Abroad	NA	50	50	0	50	NA	NA
	India	NA	10	10	90	80	NA	NA
	p-value	NA	0.72	0.72	0.07	1	NA	NA
Work status	working	NA	9.09	9.09	81.82	81.82	NA	NA
	not working	NA	100	100	0	0	NA	NA
	p-value	NA	0.35	0.35	0.54	0.54	NA	NA
Household	Living alone	NA	0	100	0	0	NA	NA
	2-4 persons	NA	9.09	9.09	81.82	81.82	NA	NA
	5-6 persons	NA	100	0	0	0	NA	NA
	more than 7 persons	NA	0	0	0	0	NA	NA
	p-value	NA	0.35	0.35	0.54	0.54	NA	NA

NA Not Applicable

3.15 Responses on getting alcohol during lockdown

A question was asked to know the modes of getting alcohol in the lockdown with ongoing restrictions. The results are presented in Table 14. It has been observed that more than one thirds (46.15%) of the population had bought enough supplies before the lockdown and were also buying on demand from supermarket or local store. However, it was also observed that another 28.85% of the respondents have been offered alcohol from friends or relatives during lockdown.

Table 14: Getting alcohol during lockdown

	N%
<i>bought enough supply</i>	46.15
<i>buying on demand</i>	46.15
<i>offered from elsewhere</i>	28.85

3.16 Attitude of respondents towards health risk behaviour post lockdown

Questions were asked to know the attitude of individuals post lockdown towards their health risk behaviour. The results are presented in Table 15. It was observed that less than half of the population (44.88%) were willing to reduce their health risk behaviour after lockdown as they faced positive outcomes on their physical and mental health and the rest (55.12%) did not consider changing their health risk behaviour.

The reasons presented by more than half (73.39%) of the population for not changing their health risk behaviour, was the habitual urge to drink or smoke. Less than one third (13.76%) of the population have been facing withdrawal symptoms in the lockdown hence they would habitually turn to their health risk behaviour once the lockdown is lifted.

Table 15: Attitude of respondents after lockdown towards risky health behaviour

<i>After lockdown</i>	N%
Yes	44.88
Reasons for no change post lockdown	N%
<i>Availability of supply post lock down</i>	8.26
<i>Facing withdrawal symptoms during lock down</i>	13.76
<i>Habitual urge to drink or smoke</i>	73.39
<i>I do not find a reason not to drink.</i>	0.92
<i>Peer pressure</i>	3.67

3.17 Association between characteristics and attitude change after lockdown

A chi- square test was carried out to know the significance of changing attitude towards health risk behaviour after lockdown. The results are presented in Table 16. More than half (57.14%) of the middle-aged population were found to be significantly interested in changing their health risk behaviour post lockdown. The population of India showed significant attitude in health risk behaviour change after the lockdown is lifted. More than half of the population not working in the lockdown (58.51%) and individuals who were not living alone were seen to have a significant positive attitude in changing their health risk behaviours due

to improved mental and physical wellbeing with reduced consumption of alcohol and cigarettes in the lockdown.

Table 16: Association between characteristics and attitude change after lockdown

		Yes%
Gender	Female	55.36
	Male	40.94
	p-value	0.06
Age group	18-27 years	37.6
	28-37 years	37.6
	38-47 years	57.14
	48-57 years	75
	58-67 years	100
	p-value	0*
Residence	Abroad	30.67
	India	53.08
	p-value	0.002*
Work status	working	33.33
	not working	58.51
	p-value	0.00*
Household	Living alone	10.87
	2-4 persons	56.83
	5-6 persons	36.84
	more than 7 persons	100
	p-value	0.0000007*

*p<0.05

4 DISCUSSION

The COVID 19 pandemic has been largely disruptive. To contain the spread of novel corona virus, the lockdown was implemented across the globe since mid-March. The lockdown however also took down on the liquor and tobacco supplying shops with their shut down. This has had an impact on the individuals who are binge-alcoholic and smokers. The approach of this study was to find out the status of such individuals and their patterns of smoking and alcohol consumption amidst the lockdown. The study protocol was adapted from online survey conducted by the health Promotion Agency on the impact of COVID 19. This study included all the individuals who practice health risk behaviour like smoking and alcohol consumption. The online questionnaire consisted of 21 questions and was circulated over the internet. It was observed that, 97.48% individuals continued smoking and 93.33% individuals continued drinking alcohol even in the lockdown with the imposed restrictions.

On further evaluation it was found that there had been discrepancies in the consumption pattern of the individuals during the lockdown. This could be attributed to several reasons which were explored through the study. Smoking was found to be practiced more than usual among individuals and this was similar to results obtained by the survey done by the Health Promotion Agency. 50.42% individuals who smoke were found to have increased their frequency of smoking in the lockdown. Result obtained by The Health Promotion Agency projected a total of 34% individuals smoking more than usual in the lockdown. The frequency of smoking more than usual in the lockdown was seen more amongst females than males. This can be attributed to the increasing trend of smoking among females.^[5] The increased smoking pattern in the lockdown was seen among individuals of age group 38-47 years, followed by individuals of age group 18-27 years. It has been found through surveys that middle aged and younger individuals have a greater smoking rate compared to others.^[6] Smoking was seen to be more among individuals living Abroad than India. Many countries in the Abroad had allowed certain small shops to be opened to supply liquor and tobacco so that people need not face withdrawal symptoms along with adapting to the negative impacts on the pandemic.^[2] Individuals working in the lockdown were seen to have an increased pattern of smoking attributing to the burn out from work leading to stress, insomnia, breathlessness eventually individuals taking on to smoke more to switch off and relax.^[7] Individuals living alone in their house have been associated with increased smoking due to the feeling of being bored in the lockdown and being at home with the restrictions of going out. However, the frequency of smoking has been observed to be reduced with individuals living with more than 4 persons in their household due to lack of privacy and unable to go out and socialise in the lockdown.

On seeing the alcohol consumption pattern, it has been less than usual in the lockdown due to the restrictions imposed on the liquor shops and difficult to get hold of alcohol. The difficulty was faced to a greater extent among individuals in India. The younger population however could not get hold of alcohol in the lockdown with the increased prices of liquor imposed on the selling.^[8] Similar to smoking pattern alcohol pattern was seen to be reduced among individuals who stay with more than 4 persons in their household.

Individuals however continued practicing their health risk behaviour as many individuals were seen to have stocked up their dose of cigarettes and alcohol before the lockdown was imposed and a few fetched their daily dose of nicotine on demand from local stores and supermarket.

Stress, anxiety, feeling of switching off or relax were seen to be the reasons perceived by individuals on taking on these health risk behaviours. Lockdown has imposed many offices and workplaces to shut down and lay off many employees to meet the economic crisis.^[8] The individuals who work from home are facing burn out from the work pressure and the ones been laid out or waiting for recruitment which has been delayed due to lockdown are anxious, stressed thinking about their financial constraints and life stability eventually impacting their mental health.^[8] However, the lockdown has also imposed a positive effect on few individuals where they have been able to socialise much with their family members and in some way, their reduced consumption patterns have had a positive impact on their physical and mental health eventually persuading them to think of changing their habits post lockdown.

5 CONCLUSION

Lockdown imposed to contain the spread of COVID 19 had it's positive and negative impacts on individuals. Where a few individual perceived their change in behaviour habits to have brought in a positive impact on their mental and physical health and are looking forward to quit their habits hereon, there are individuals who cannot deprive themselves of their daily dose of nicotine and lockdown imposing them with withdrawal symptoms. Pressure from peers and habitual urge to consume alcohol and smoke with their availability post lockdown cannot restrict them from practicing their health risk behaviour. The psychological code of conduct plays a major role in driving individuals towards health risk behaviour. The lockdown has turned to be constructive to the mental health and decreased pattern of smoking and consuming alcohol for a few individuals and for other it has turned into being destructive.

6. REFERENCES

1. Chutel, Lynsey. Taking on Covid-19, South Africa Goes After Cigarettes and Booze, Too. *The New York Times* . 8 May 2020.
2. Naming the coronavirus disease (COVID-19) and the virus that causes it. *World Health Organisation*. February 2020.
3. Tandon, Suneera. Covid-19 to impact out-of-home alcohol consumption. *The Live Mint*. 12 May 2020.
4. Mishra, Subhangi. How smokers are getting by without cigarettes in Covid-19 lockdown (it's helping many quit). *The Print* . 26 April 2020.
5. Gender And Tobacco . *World Health Organisation*. 2010.
6. Leading Health Indicator - Tobacco . *HealthyPeople.gov*. 2020.
7. Reuters. Lockdown sends grey-market booze prices soaring. *The Economic Times - Retail*. 15 April 2020.
8. Neilson. The Impact of COVID 19 . *Health Promotion Agency* . 17 April 2020.
9. Work from home burnout: Stress, insomnia, anxiety becoming the new normal. *The Economic Times*. 2020.
10. Derek Yach, DSc (Hon), MPH. Tobacco Use Patterns in Five Countries During the COVID-19 Lockdown. *Nicotine and Tobacco Research* . 2020.
11. Bureu, EH News. 68 per cent smokers reduced cigarette consumption during COVID-19 lockdown: Indus Health Plus survey. *Express Health Care* . June 1, 2020 .
12. Kiely, Melita. Brits drinking less alcohol during lockdown. *The Spirits Business* . 5 June, 2020.

7. ANNEXURE

Online questionnaire form -

<https://docs.google.com/forms/d/16L28P71gcoOilTJW3jPUprmFzJpWeysWVRRqYb6tI/edit>

SUMMER TRAINING REPORT

PART B

DATA ANALYTICS IN LEAN SIX SIGMA

A Report

By

Preetha P

Under the guidance of

Dr Arindam Das

MBA Hospital and Health

Management 2019-2021

 **IIHMR UNIVERSITY**
Institute of health management research
The IIHMR University,
Jaipur 2019

ACKNOWLEDGEMENTS

I would like to extend my true appreciation to Dr P.R Sodani, Pro-President Dean, Training Dean, Institute of Health Management Research for proposing this seminar on Data Analytics in Lean Six Sigma on Coursera platform.

I am appreciative to my guide Dr Arindam Das, Associate Professor, Institute of Health Management Research for the consistent help identified with my questions, for his understanding, inspiration, and enormous information. His direction helped me in all the time recorded as a hard copy of this report.

Last yet not the least, I offer my thanks to Dr Inez Zwetsloot, the course teacher who is an excited scholarly scientist just as a down to earth business expert in data analytics and tasks the executives at the Amsterdam Business School. In 2016, she effectively guarded her PhD in mechanical measurements at the University of Amsterdam. Close to her exploration she has encouraged Lean Six Sigma courses to both experts and understudies in The Netherlands, since 2012.

TABLE OF CONTENTS

S.no	Content	Page number
1	Introduction	41
2	Improving process through Lean Six Sigma	42
3	CTQ- Critical To Quality	43
4	Units and Operational Definition	43
5	Random sampling using Minitab	44
6	Organising data	45
7	Numerical and categorical data	46
8	Descriptive Statistics using Minitab	46
9	Visualising numerical data	48
10	Visualising categorical data	50
11	Pareto analysis	51
12	Visualising two variables	52
13	Estimation and confidence intervals	54
14	Normal, Lognormal and Weibull distribution	55
15	Probability plot	56
16	Empirical CDF	57
17	Data based testing in Lean Six Sigma	58
18	Hypothesis testing	58
19	Causality	59
20	ANOVA Analysis	59
21	Kruskal- Wallis Analysis	62
22	Two-sample t- test	62
23	Test for equality of variances	63
24	Correlation	65

25	Regression analysis	66
26	Quadratic regression	71
27	Chi-square analysis	71
28	Logistic regression	73
29	Conclusion	78
30	References	78

LIST OF FIGURES

S.no	Content	Page number
1	Random sampling using Minitab	43
2	Random sampling using Minitab	44
3	Random sampling using Minitab	44
4	Numerical and categorical data	45
5	Displaying descriptive statistics using Minitab	46
6	Descriptive statistics- N,N*, Mean, Standard deviation	46
7	Descriptive statistics- Range, Quartiles	47
8	Visualising data	47
9	Histogram- Bulk, tails	48
10	Histogram distribution	48
11	Box plot interpretation	49
12	Pie chart, bar graph	50
13	Tally table	50
14	Pareto analysis	51
15	Visualising two variables	51
16	Scatter plot diagram	52
17	Stacked bar chart	53
18	Estimation of population	53
19	Estimators	54
20	Normal distribution	54
21	Properties of normal distribution	54
22	Normal distribution- largest location	55
23	Normal distribution- largest spread	55

24	Rules of thumb – Normal distribution	56
25	Weibull distribution	56
26	Lognormal distribution	56
27	Probability plot	57
28	Empirical CDF – Example	58
29	Data analysis- suitable method	58
30	ANOVA- Hypothesis testing	59
31	ANOVA – R squared	60
32	ANOVA – Residual analysis	61
33	ANOVA – Residual analysis	61
34	ANOVA – Residual analysis	62
35	Kruskal- Wallis analysis	63
36	Two sample t-test	64
37	Test for equality of variances	65
38	Positive correlation	66
39	Negative correlation	66
40	Strong/weal positive correlation	67
41	No correlation	67
42	Regression analysis	68
43	p-value and R squared	69
44	p- value and R squared – examples	71
45	Regression- residual	71
46	Regression- residual analysis	72
47	Prediction interval	72
48	Chi- square output – Count	73
49	Chi- square output- Row percentage	74

50	Chi- square output- Column percentage	74
51	Chi-square p-value	74
52	Event trial format	75
53	Binary fitted line plot	76
54	Binary fitted line plot- p- value	77

LIST OF ABBREVIATIONS

DMAIC	Define, Measure, Analyse, Improve, Control
Ph.D.	Doctor of Philosophy
CTQ	Critical to Quality
CDF	Cumulative distribution function
P-value	Probability value
R-squared	Regression- squared
Q1	Quartile 1
Q3	Quartile 3
IQR	Interquartile range
THT	Total handling time
AD statistics	Anderson darling statistics
ANOVA	Analysis of Variance
UvA	University of Amsterdam

SUMMARY

The course presents Lean Six Sigma which is a technique that depends on a synergistic collaboration to improve performance and shows where data and data analytics have their place inside the DMAIC system. It additionally presents the product bundle Minitab (a statistical programming software for data analytics). The necessary pledge to the course was 5 weeks (4 hours of study for each week) covering an aggregate of 20 hours of study time. The course was started on April 20,2020 and completed on May 15,2020. The course incorporated video addresses, readings, practice works out, conversation discussion and reviewed tests toward the finish of the module. The first week covered the utilization of information in Lean Six Sigma undertakings and how to compose information. A prologue to Minitab, the statistical software which is utilized for data analytics was given. Further talks spread the ideas on comprehension and envisioning information utilizing diagrams like histogram, box plot, bar outlines, pie outlines, pareto outlines with translation of these charts and besides, required to summarise data. A module on probability distribution was shrouded in the second week with a target to quantify uncertainties and responding to business questions. The third week module was on testing, instructed to formulate statements that can be done using data analytics. After this, two modules on statistical testing, in which methods to establish relationships between variables was taught using correlation and regression techniques. The arrangement of video addresses in the course was about how to utilise data analytics for functional issues at a hierarchical level and had a solid spotlight on the use of data analytics inside the Lean Six Sigma structure.

I. INTRODUCTION

Seeing how to interpret and utilise data is a significant expertise. An aptitude that is turning out to be increasingly more helpful nowadays, as the measure of accessible information is expanding. Lean Six Sigma is a procedure for process improvement, and it utilises data and data analytics.

Questions that can be examined inside a Lean Six Sigma venture are:

- What is the most significant issue that an association should concentrate on?
- How to evaluate a specific marvel?
- Do the workers impact on what amount time it takes to deal with an occupation?
- Which division shows the best performance?
- Which division is the most reliable, or shows the highest calibre?
- How to diminish the quantity of deformities?
- Is there a genuine distinction in quality in a venture?

These kinds of questions are addressed utilising data analytics in a Lean Six Sigma venture. The utilisation of data analytics apparatuses has, truth be told, a lot more extensive application zone. Data analytics can be valuable in different regions where one experiences information. The key ideas canvassed under Data Analysis in Lean Six Sigma are as per the following:

Week 1

- Introduction to Lean Six Sigma
- The five phases of DMAIC
- Introduction to CTQ measurement
- Collection of a representative sample
- Organization of data and distinguish between units and variables
- Operational definition corresponding to a CTQ

Week 2

- Numerical and categorical data
- Descriptive statistics
- Histogram and boxplot interpretation
- Bar graph, pie diagram, count table interpretation
- Pareto graph interpretation
- Creating an appropriate graphical visualisation for two variables and interpreting the results

Week 3

- To comprehend the connection between a sample and population
- To comprehend the concept and use of estimation
- To quantify precision of estimate (using confidence intervals)

- To distinguish various kinds of probability distributions (Normal, Lognormal, Weibull)
- To determine the distribution of numerical variation using probability plot
- To determine the probabilities and percentages using Empirical CDF
- To use "rule of thumb" to compute probabilities for normally distributed variables
- To comprehend parameters of normal distribution

Week 4

- To recognise X variable and Y variable in practice and select the suitable test
- To interpret p-value of hypothesis testing
- To interpret results accurately by distinguishing between causality and relation

Week 5

- To interpret and use correlation
- To judge when regression is the pertinent technique
- To play out a regression and interpret output, particularly p-value and R-squared
- To process and interpret a prediction interval
- To perform and interpret a quadratic regression
- To judge the validity of regression using residual analysis

II. IMPROVING PROCESS THROUGH LEAN SIX SIGMA

Lean six Sigma is a data driven way to deal with process improvement. Processes are progressions of steps made to make an item or product a service. Processes occur all over the place. For instance, in industry. One model is embodiment process of medication. Here a medication is placed into containers. A regular procedure in social insurance is the treatment of the patient in a clinic. There are additionally numerous procedures and administrations. Each association can be viewed as an assortment of procedures. What is more, numerous associations invest significant energy in improving these processes to withstand rivalry, to adjust to higher client desires, and furthermore enhance mechanical advancement.

Improvement on account of the exemplification procedure may intend to deliver more containers every hour so deals can increment. In emergency clinics a potential improvement bearing is to diminish the event of clinical mistakes. Also, in a bring focus enhancements over procedure can intend to build consumer loyalty by bringing down holding up times.

Be that as it may, how might one approach be improving process?

An extremely well-known approach to improve process is Lean Six Sigma. This is a stepwise strategy for running a process improvement venture. So, principle highlights of Lean Six Sigma are its proof based and

that it utilises process information. Information can be utilised in a task to investigate a few parts of a procedure. For example, one can utilise the information to locate the most serious issues that one should concentrate on. Information is significant when one needs to improve forms. Data can assist one with finding the most serious issues, analyse the current procedures or to explore the impacts of progress activities.

Lean Six Sigma offers a guide known as DMAIC or D-M-A- I-C. These five stages endorse a stage shrewd technique for process improvement dependent on estimations, and proof-based improvement activities. Thus, data and data analysis are significant in any DMAIC venture.

The D represents Define, the M represents Measure, A for Analyse, I for Improve, and C for Control. An individual who drives a Six Sigma Project is known as a belt.

In the Define stage the belt chooses a task and builds up venture goals.

In the Measure stage the belts make issue quantifiable and measures the procedure execution. After the measure stage comes the break down stage, wherein the belt examinations the current circumstance and makes an analysis for diagnosis. The objective is to decide the idea of the current issue. After finding of the current procedure a belt creates and executes improvement activities dependent on fundamental impact factors in the improve stage. At long last in the control stage, the procedure controls are improved, and the undertaking is shut.

III. CTQ –CRITICAL TO QUALITY

In the DMAIC stages, the piece of CTQ goes under the measure stage, which is tied in with making an issue quantifiable and quantifiable. This is likewise called the operationalization of an issue.

One needs to have process measurements, that are important for venture destinations. Or then again, as such, what qualities would it be advisable for you to gauge?

A CTQ is a property of an item or an assistance that is applicable to extend objective. One can discover it by asking, what is the metric one needs to change? In Lean Six Sigma phrasing, this measurement is known as a CTQ. In different fields, be that as it may, this is regularly called Y-variable, dependent variable, or the outcome variable.

IV. UNITS AND OPERATIONAL DEFINITION

Units are individuals, articles, or phenomenon. Units are the things one gathers information on. The estimation unit characterises what an estimation implies. The measurement units are the properties of the things one gathers information on. The concession to how to gauge a variable and what the unit is called an operational definition. An operational definition can be minimal self-assertive, it can change dependent on your procedure of intrigue however should be a reasonable one.

A guide to delineate the idea of unit, estimation unit and activity definition:

One is keen on finding the first run through fix rate in a call place (that is, the quantity of calls that a representative can deal with without connecting a client with another division)

So the Y-variable, CTQ, is first time fix rate.

What is the unit? The appropriate response is the information one gathers. For all the calls one gathers, yes or no?

Yes, means the worker dealt with the call. What is more, no, which means the worker needed to connect the call with another person.

In any case, one is keen on the first run through fix rate, the level of calls that are taken care of the first run through. This must be estimated in a time span, for example, every day.

The unit is per representative every day. Or on the other hand it could likewise be per worker every week (which makes the operational definition).

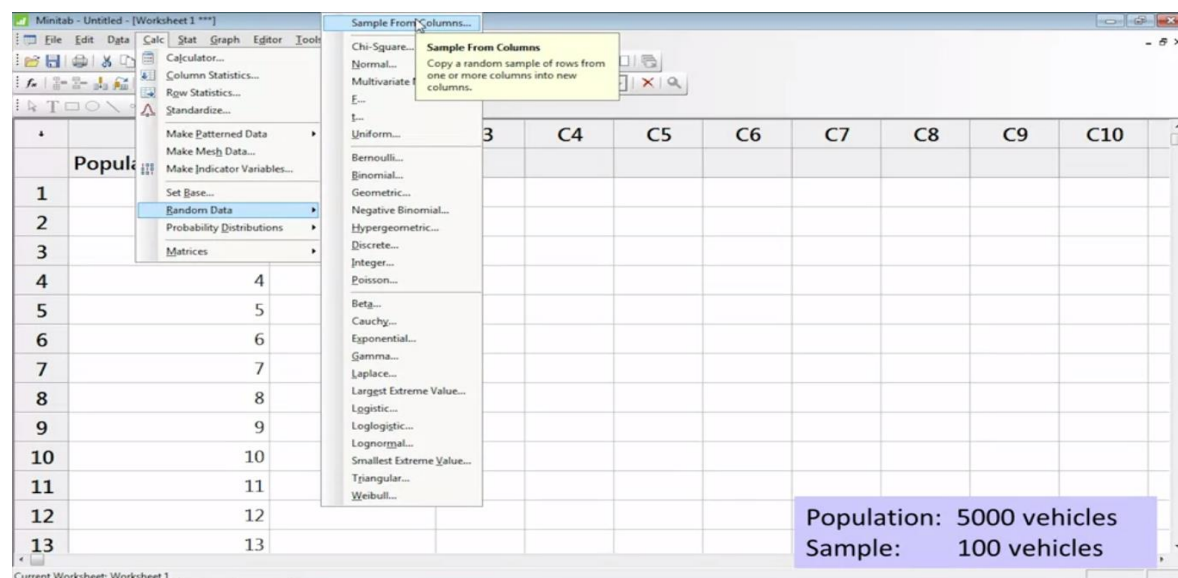
For every day, or week, and for every worker, one estimates a level of calls that have an indeed, and this give the first run through fix rate. The measurement unit is the level of calls that are dealt with first time right.

V. RANDOM SAMPLING USING MINITAB

How can one gather a sample? What makes a sample a decent example?

Indeed, a sample must be representative. Actually, a representative sample implies that the sample ought to be founded on a system that is free of the inquiry that one needs to reply. Determination of the sample ought to be done haphazardly. One ought to consistently request that a PC do this as people are not modified to make arbitrary determination. This should be possible utilising Minitab statistical programming.

Making a random choice of 100 vehicles from a populace of 5000 vehicles



*Figure 1: Make a section in the Population vehicles segment beginning with 1 and as far as possible tallying up to 5,000. To choose an arbitrary example of 100 vehicles go to **Calc**, at that point to **Random Data** on the grounds that an irregular choice is required and go to the choice **Sample** from a Column.*

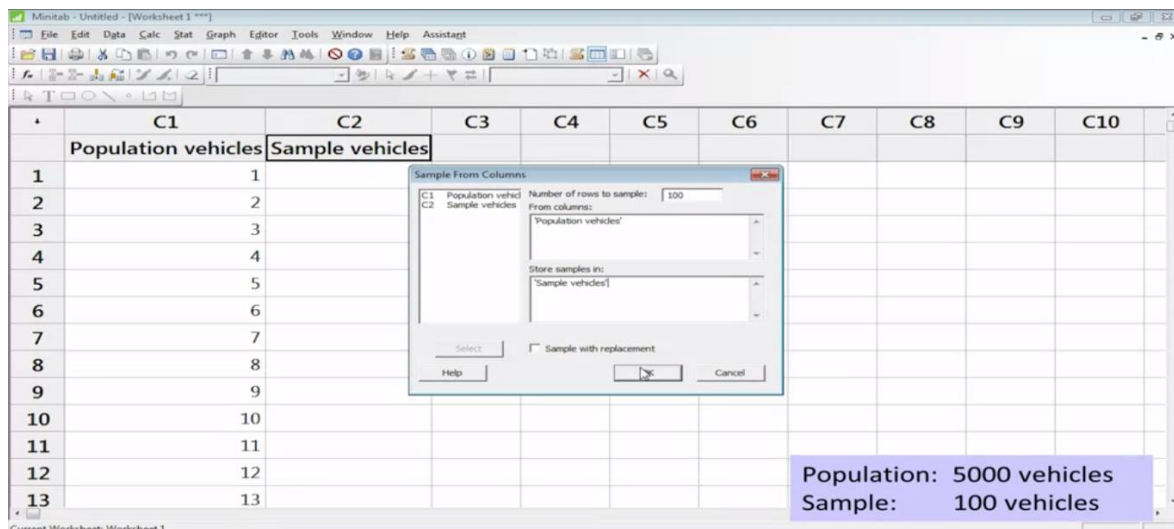


Figure 2: To get an example of size 100. Sample this from the section Population vehicles and store this in the segment Sample vehicles, enter okay.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	Population vehicles	Sample vehicles								
1	1	492								
2	2	2667								
3	3	3440								
4	4	4076								
5	5	4720								
6	6	2229								
7	7	372								
8	8	2529								
9	9	4562								
10	10	3127								
11	11	2363								
12	12	3715								
13	13	38								

Figure 3: Minitab would randomly select 100 vehicles from a populace of 5000 vehicles. As we can see in the figure, the first sample vehicle selected as 492.

VI. ORGANISING DATA

The fundamental method to compose data is in a structure dependent on units and variables. A unit speaks to an individual case, what is being estimated. The variables are the properties of the units that are being measured. Units are constantly put away in the columns of data set. A unit is known as the experimental or observational unit or at times alluded to as a case. A data set additionally comprises of variables, which is constantly put away in the rows. A variable speaks to the properties of the units estimated. Sorting out information makes investigations of information simpler.

VII. NUMERICAL AND CATEGORICAL DATA

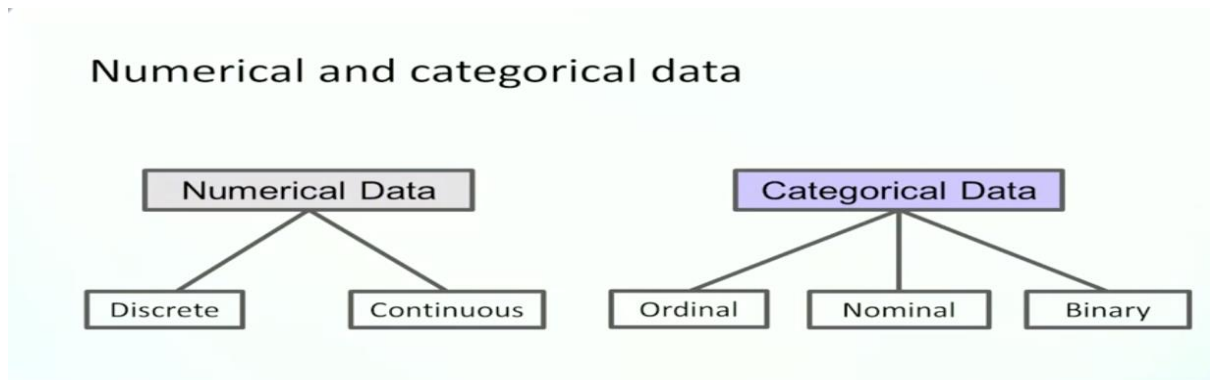


Figure 4: Numerical and Categorical data

There are two principle classifications of data. The first is numerical, and the subsequent one is continuous. On the off chance that one spotlights on numerical, a development among discrete and continuous data is made. For straight out, one can make a development between ordinal, nominal, and binary data.

Numerical data is a quantitative data. The data depends on numbers that one can ascertain with, for example, temperature, dimensions, or the quantity of individuals in a room. One can differentiate between discrete, which is check information, or information that is adjusted, and with variables that are continuous, like temperature, that can be any sort of number.

Categorical data is qualitative information. Thus, one can divide the information into groups. On the off chance that there is a significant request to the groups, it is ordinal information. If the groups comprise of unordered classes, it is nominal information. If one just has two alternatives accessible, we will call it binary information

VIII. DESCRIPTIVE STATISTICS USING MINITAB

Enlightening data using Descriptive statistics is just accessible for numerical data since they depend on mathematical operations like addition and multiplication.

One can approach Minitab for Descriptive statistics under the Stat menu, under Basic Statistics. Afterwards, the primary choice is Display Descriptive Statistics.

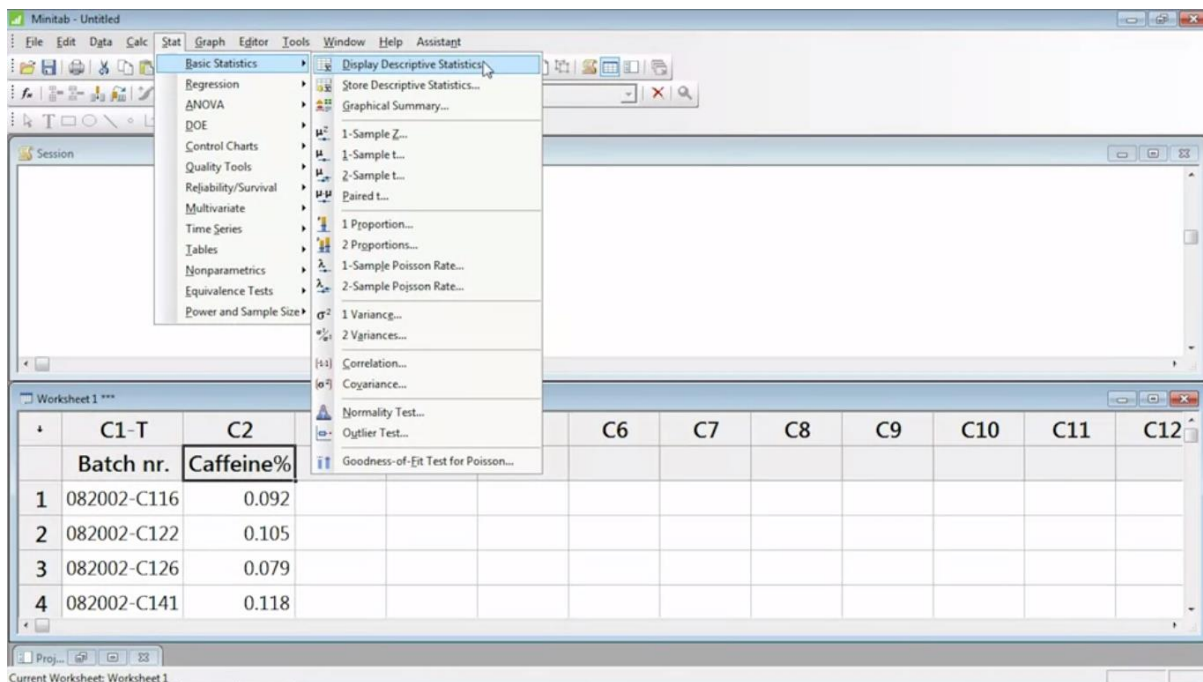


Figure 5: Displaying descriptive statistics in Minitab. The information, with batch number in one column (C1-T), and Caffeine% (C2) in the other

In the figure below, Minitab reports N and N*. N shows the number of measurements in data, and N* demonstrates the number of missing values in data. 40 values are accessible, and 0 are missing in the figure beneath. Mean, which is likewise called the average, is the total of all estimations partitioned by N and it is frequently used to demonstrate the location of the data. There is consistently a specific level of variation or scattering in information. A low measure of variation implies that most failures are somewhat near the mean. A high measure of scattering implies that most failures are very far away from your mean and spread out over a bigger span. To gauge scattering one needs to take a gander at the standard deviation.

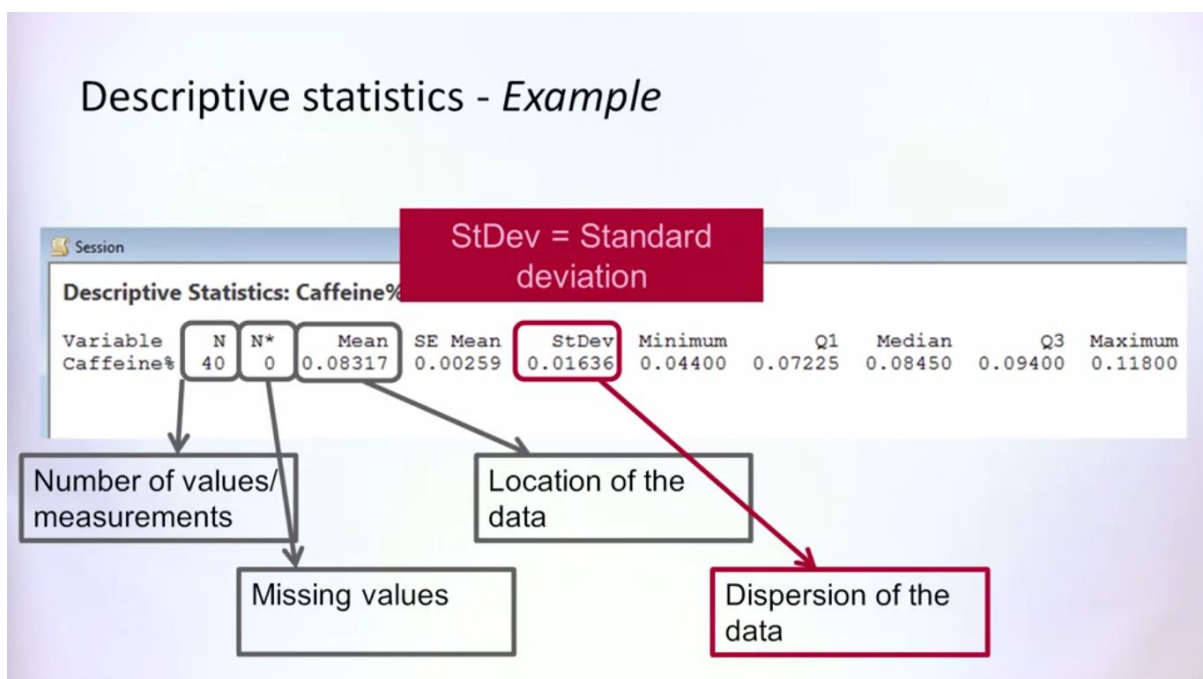


Figure 6: Descriptive statistics using Minitab- N, N* Mean, Standard deviation

Much the same as the mean, the median is a measure of central tendency. Contrasted with the mean, it has the preferred position that it is not influenced by anomalies or outliers. In any case, the weakness is that it is not utilising all information focuses. It just uses the middle value. Also, it is subsequently less exact than the mean.

Q1 and Q3 mean the first and the third quartile.

Q1 is that value that isolates the least 25% of the information from the most elevated 75% of the information. Q3, the third quartile isolates the least 75% of the information from the most elevated 25% of the information. The separation somewhere in the range of Q1 and Q3 is frequently used to evaluate dispersion and is known as the interquartile range. The interquartile range is less susceptible to the impact of outliers.

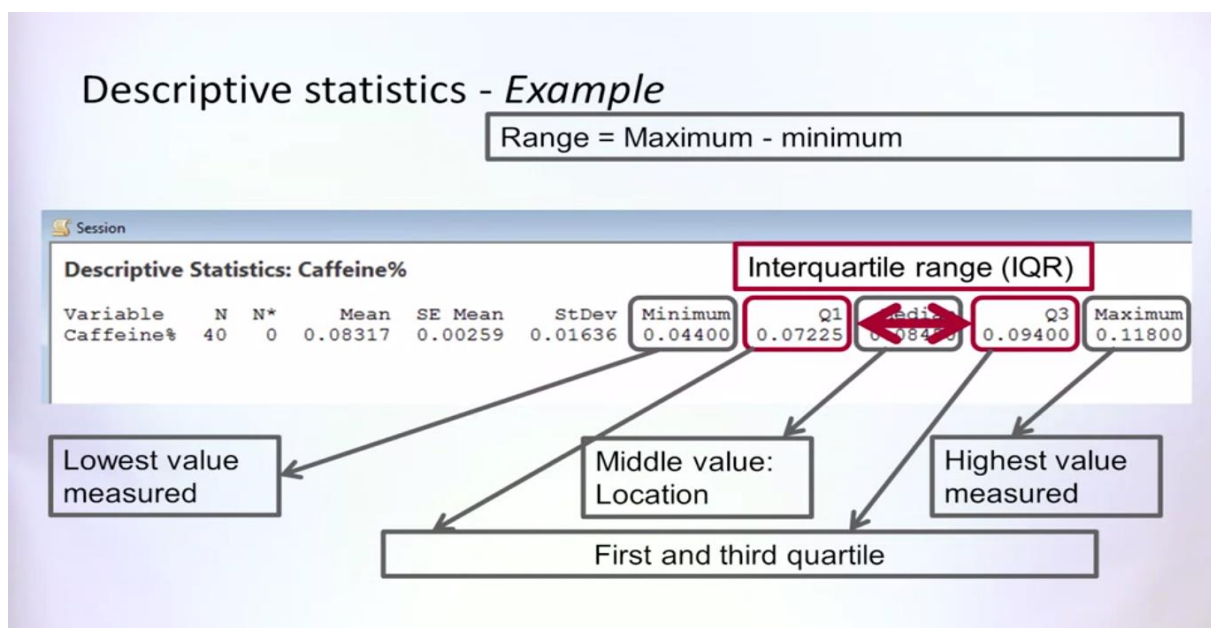


Figure 7: Descriptive statistics using Minitab- Quartiles, Range, Interquartile range

IX. VISUALIZING NUMERICAL DATA

For a numerical variable, the histogram and the boxplot are the most mainstream techniques for representation of this variable. To make a histogram or a boxplot in Minitab, one needs to go to the menu Graph and will discover histogram and box plot.

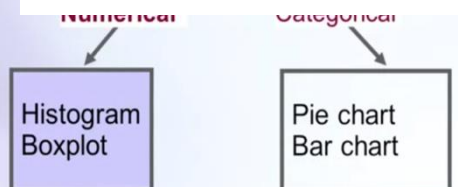


Figure 8: Visualising data

Histogram: The bigger piece of the information that is thought together is known as the bulk and the rest of the parts on the two sides are known as the tails. Tails decline outwards.

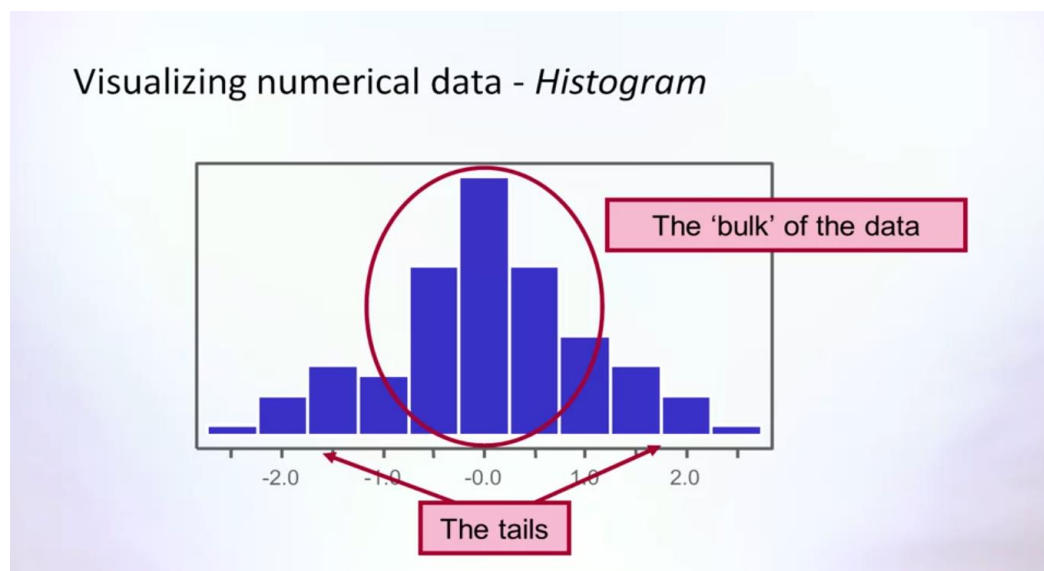


Figure9: Histogram- Bulk, Tail

Histograms can be utilised to survey fascinating examples with regards to data. It very well may be utilised to see whether information is symmetrical or skewed, or bimodal. Bimodality regularly happens when two particular populaces are available in the example, giving a blend of circulations.

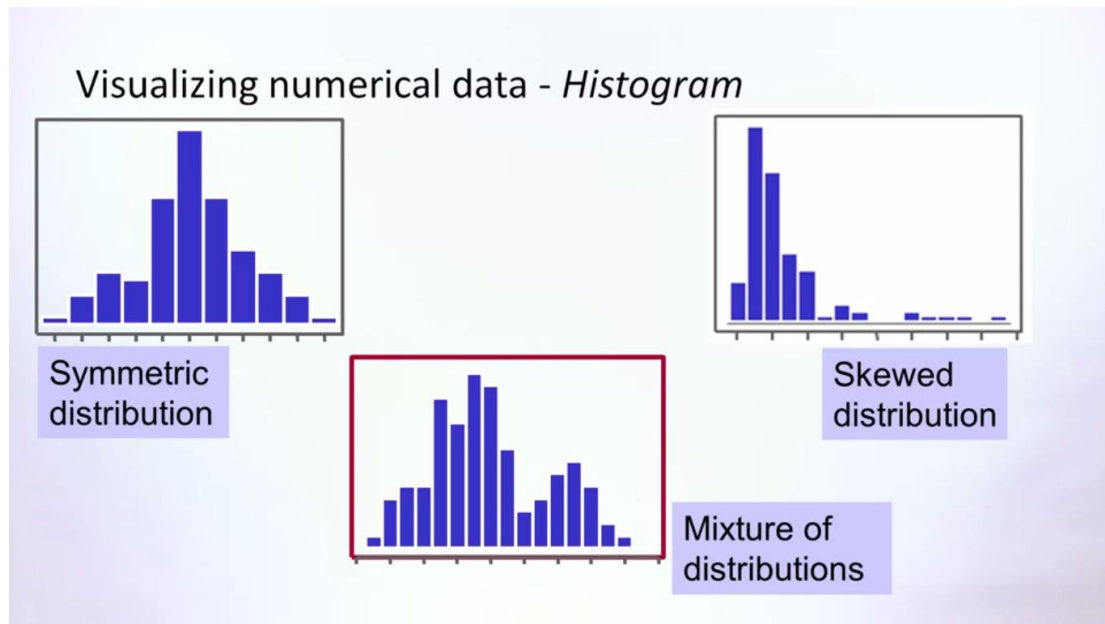


Figure 10: Histogram distribution

Box Plot:

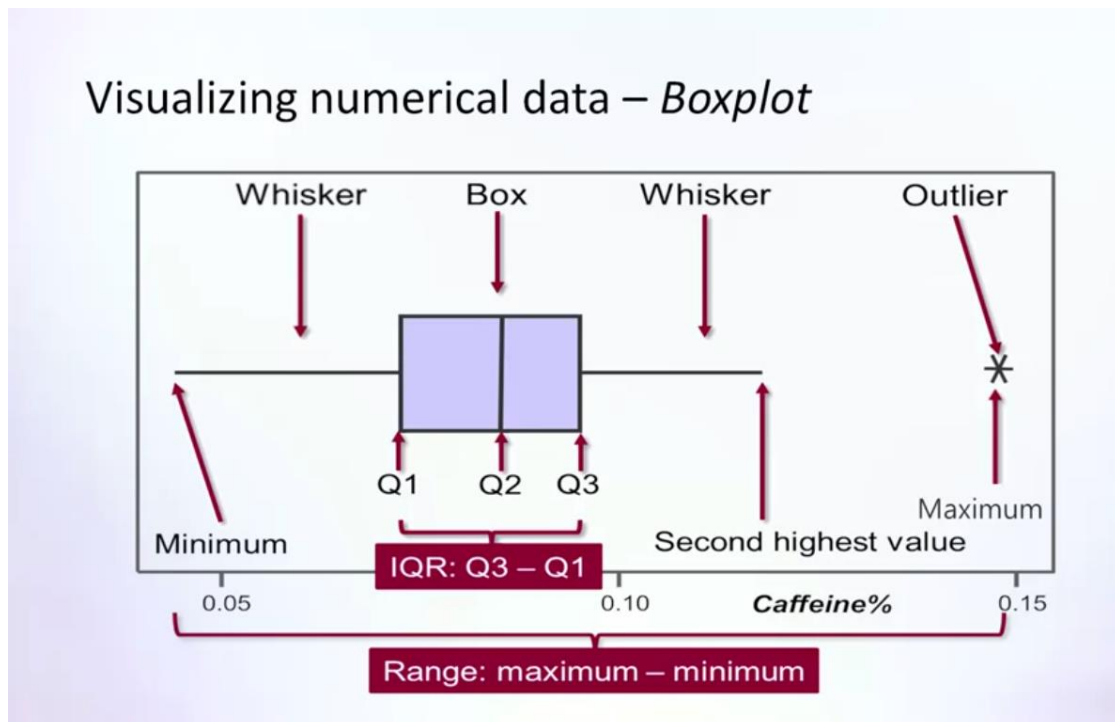


Figure 11: Box plot interpretation

A boxplot comprises of a box which is the centre half of information, that is 25% of the information is littler and 25% of the information is bigger. These are shown by the whiskers. If there are perceptions that are far away from the case, these perceptions are called outliers and are shown with a star.

A boxplot uncovers where the information is situated on the genuine line. It shows you the base worth and the most extreme worth. On the off chance that there is one anomaly, as in this figure above, at that point the second most elevated worth figures out where the whisker will stop. The boxplot additionally shows the interquartile range. The interquartile range is equivalent to the length of the box. What is more, that is equivalent to the difference between the third quartile and the first quartile. The boxplot shows the total range is equivalent to the difference of the most extreme value and the least value.

X. VISUALISING CATEGORICAL DATA

For a single categorical value, a pie graph, bar outline or a count table can be utilised. A pie diagram shows how regularly every classification happens. A bar diagram gives a similar data; however, it is simpler to spot which class happens regularly. A count table gives the specific numbers.

Minitab helps in envisioning categorical variables with the assistance of pie diagrams and bar graphs, this should be possible by setting off to the Graph menu and picking Pie outline/bar graph presentation.

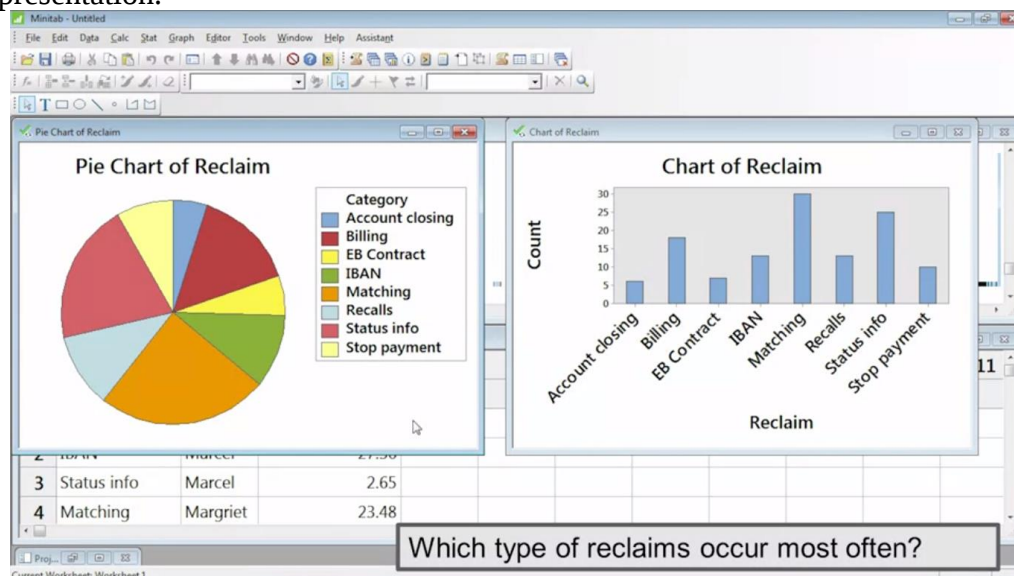


Figure 12: Pie chart, Bar graph

Tally table: To make a tally table, in Minitab statistical software, go to Stat, and then to the menu Tables and choose the first option, the Tally Individual Variables. A tally table gives the exact numbers and percentage of the required data.

Visualizing categorical data – Tally table

Session		
Tally for Discrete Variables: Person		
Person	Count	Percent
Henk	17	13.93
Jan jr.	17	13.93
Jan sr.	14	11.48
Karel	15	12.30
Kees	15	12.30
Marcel	22	18.03
Margriet	22	18.03
N=	122	

Figure 13: Tally table

XI. PARETO ANALYSIS

80% of all issues can be ascribed to 20% of the causes. The horizontal axis contains the kind of issues. The tallness of the bars in the diagram shows how regularly these issues happen. The bars are arranged in diminishing frequencies, with the end goal that the most every now and again happening issues are put to one side, and the numerous minor issues are set to one side. The frequencies can be perused from the vertical hub on the left.

Pareto graph with the frequencies just shows how regularly every issue happens. Be that as it may, often one is increasingly inspired by which issue occupies the vast majority of time. In such cases one ought to likewise quantify the duration.

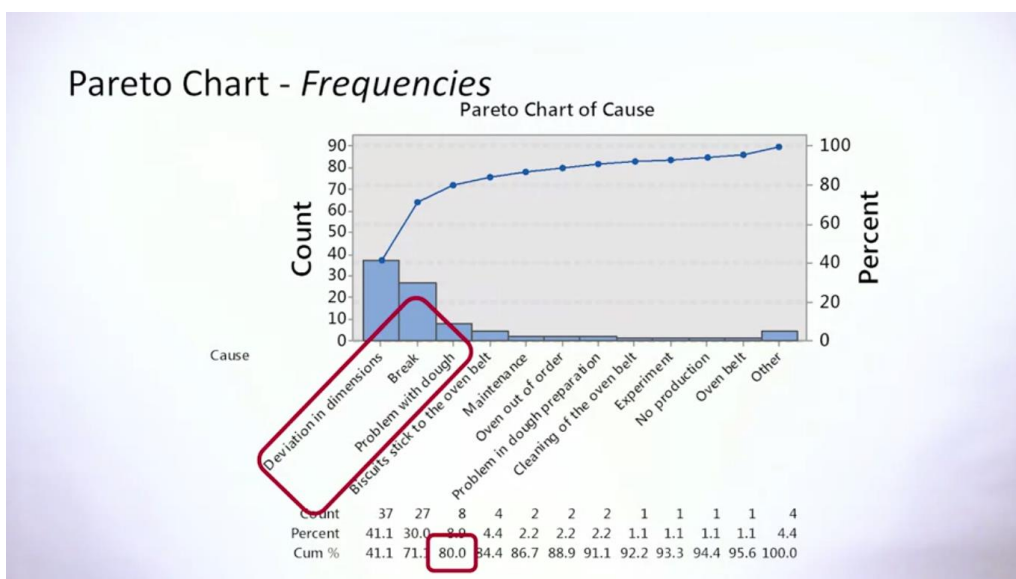


Figure 14: Pareto analysis

XII. VISUALISING TWO VARIABLES

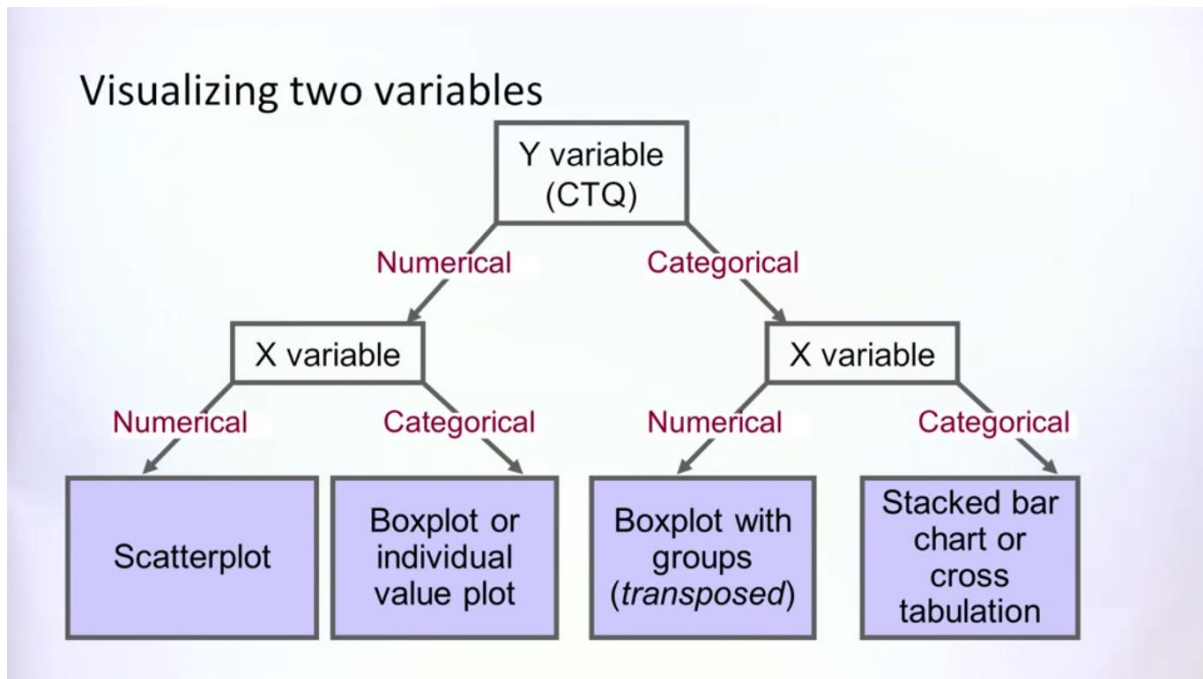


Figure 15: Visualising two variables

Scatter plot: In figure 16, beginning from the upper right corner, the dots are grouped together and pretty much structure a straight line that goes down. Along these lines this shows a straight relationship with an outlier. In the following plot, there's marginally curved line of dots. This is demonstrated of a curved relationship. In the following chart, the data points are grouped. Also, in the last Scatterplot, information is spread all around and accordingly, the variables have no relationship.

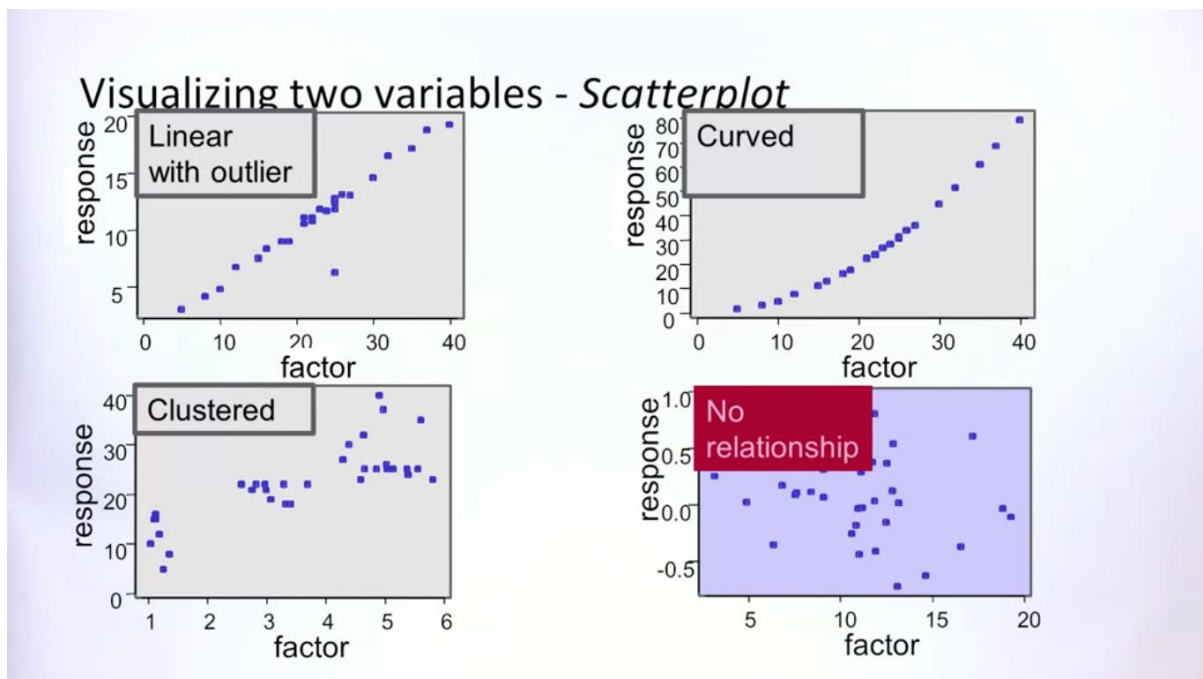


Figure 16: Scatterplot diagram

At the point when the Y variable is numerical and the X variable is categorical, a box plot is utilised for perception. Allude figure 11 for box plot.

A stacked bar graph is utilised when the Y variable and the X variable are categorical. Figure 17 delineates a stacked bar outline where the sort of specialist in taken on the horizontal axis and the count is taken on the vertical axis. Specialist P rewarded around 100 patients. Around 80 of them were in division A1, and the others were in office A4.

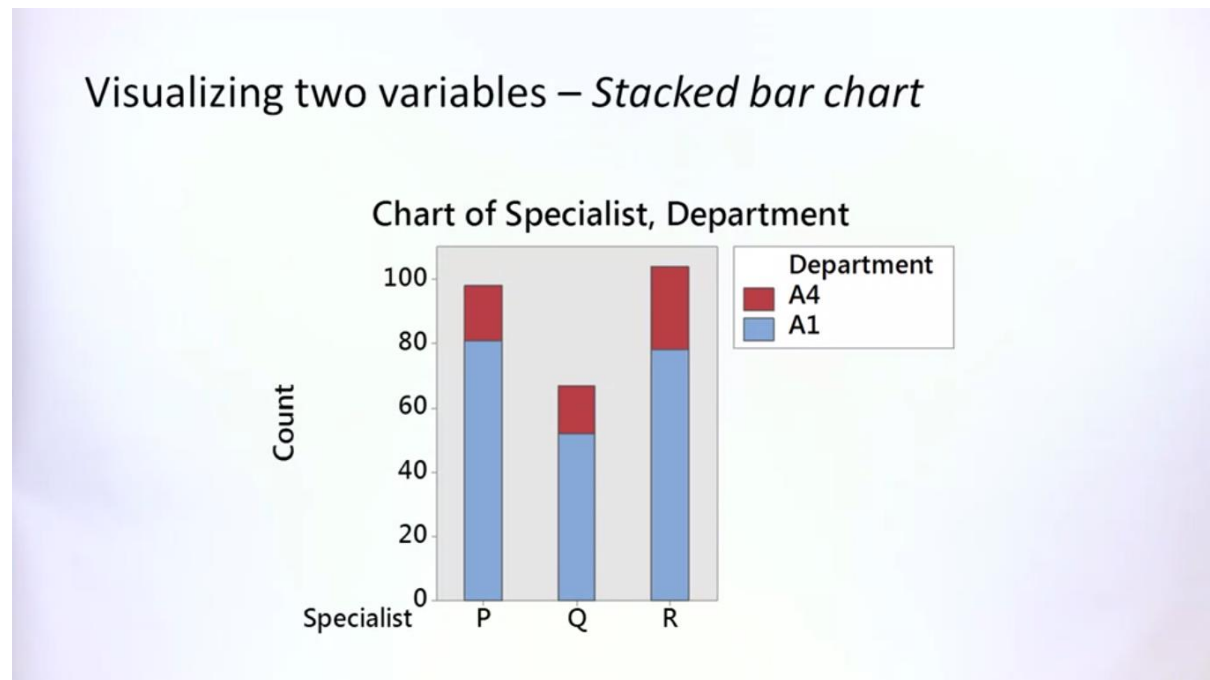


Figure 17: Stacked bar chart

XIII. ESTIMATION AND CONFIDENCE INTERVAL

Estimation is something one does frequently. What financial plan would it be advisable for you to put something aside for excursion? To what extent will it take you to get the opportunity to work in heavy traffic? Or then again does one have enough gas to return home? These are for the most part instances of estimations.

In figure 18 beneath, the blue bars are the histogram of 40 values in this sample. The red line is the estimated, or fitted, populace distribution. One ought to be intrigued to recognise what this conveyance resembles in the feeling of where its area is and what the spread is. The area is signified by the Greek letter mu. What is more, the spread is signified by the Greek letter sigma

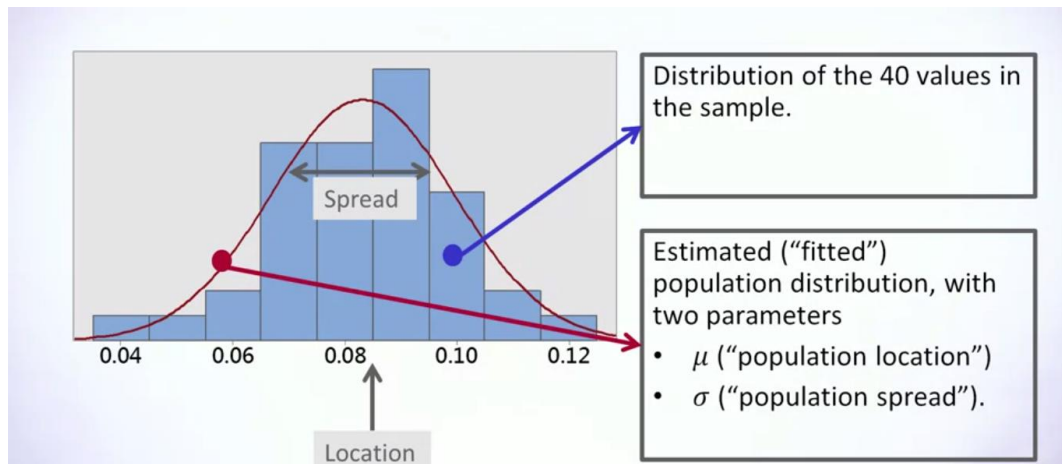


Figure 18: Estimation of a population

For evaluating the area, utilise the mean. What is more, another option is the median. So which estimator to utilise? It relies upon the circumstance. The mean is increasingly exact, as it utilises all information, yet the middle is progressively vigorous to exceptions. So, both can be helpful. For sigma, the spread in the information, the regularly utilised estimator is the example standard deviation.

However, one can likewise utilise the interquartile extend, which is a contrast between Q3, the third quartile and Q1, the first quartile. Or on the other hand one can utilise the range, the distinction between the most extreme and least. Once more, it relies upon the circumstance, which estimator is generally valuable. The example standard deviation is exact as it utilises all information esteems in the example. Notwithstanding, much the same as the range, it very well may be delicate to exceptions. The interquartile range is increasingly vigorous to exceptions.

Estimators for	
Location (μ)	Spread (σ)
Average (mean) $\bar{X}$	Standard deviation S
Median	Interquartile range $Q_3 - Q_1$
	Range $max - min$

Figure 19: Estimators

All things considered; statistics is tied in with making inferences under vulnerability. What is more, to evaluate this vulnerability in gauges, one can compute two limits, in the middle of which they are genuinely sure that the real populace mean will lie. Also, this is known as the confidence interval.

Furthermore, the regularly chosen certainty level is 95%. Minitab registers such a confidence interval. A certainty stretch gives limits in which a population boundary will lie with a specific degree of certainty.

XIV. NORMAL, LOGNORMAL AND WEIBULL DISTRIBUTION

In a Lean Six Sigma venture, one generally experiences a Normal, Lognormal and Weibull distributions.

The Normal distribution is even or symmetrical and it looks somewhat like a bell shape. The blue bars in the histogram show the circulation of 50 values in an estimated sample. The red curve is the population dispersion that Minitab fitted to the information. The more N builds up, the more it begins to appear as though a bell shape.

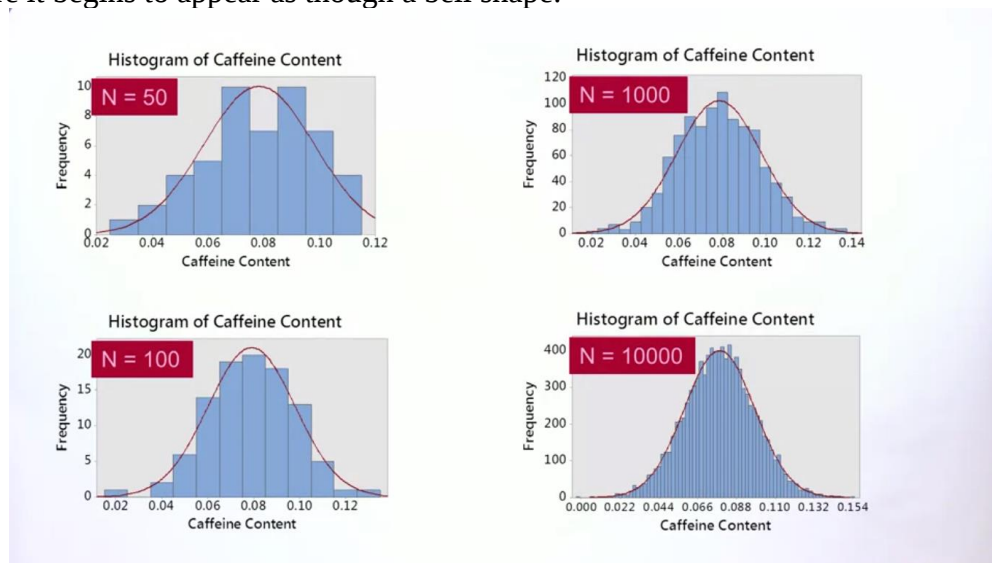


Figure 20: Normal distribution

A Normal Distribution has two parameters that decide the shape of its distribution.

The principal parameter is the location parameter which decides the midpoint. This is indicated by μ , which is the location or mean. The subsequent parameter portrays the state of the distribution and is known as the dispersion parameter, indicated by σ . Also, it decides the spread of the distribution.

Properties of the Normal Distribution

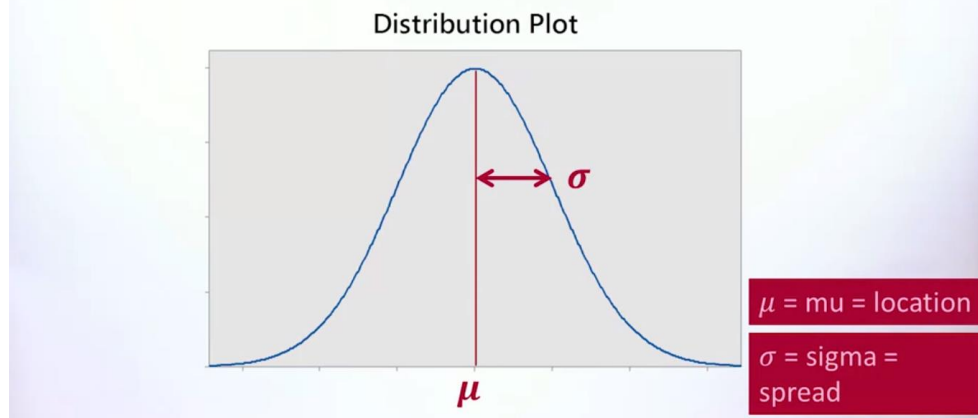


Figure 21: Properties of Normal Distribution

Properties of the Normal Distribution

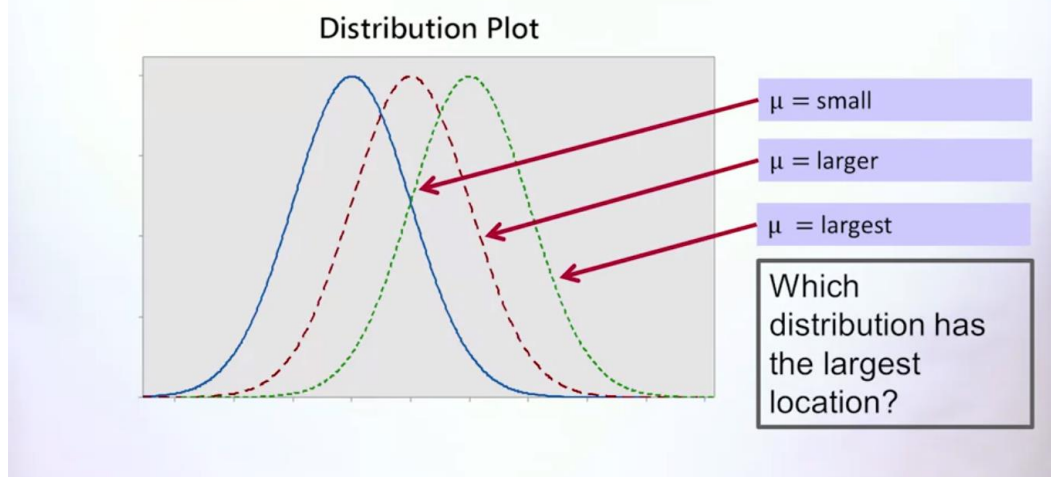


Figure 22: Normal distribution- largest location

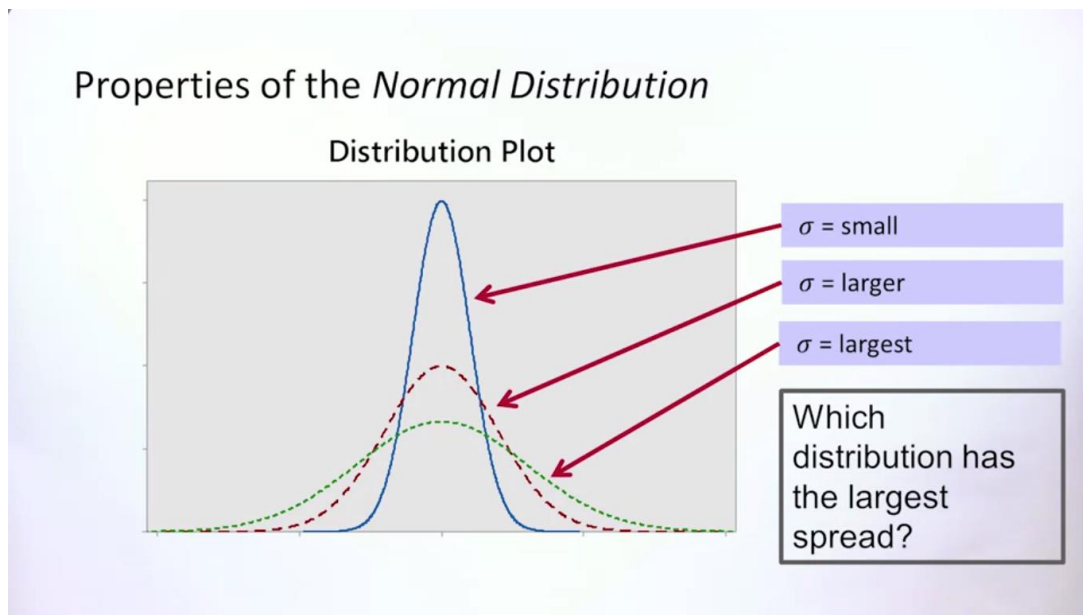


Figure 23: Normal distribution- largest spread

A variable that is normally distributed will consistently be, for 68% of the time, in the middle of the new μ in addition to 1 sigma and less 1 sigma. Proportionally, 95% of the populace will be in the middle of μ in addition to 2 sigma and less 2 sigma. 99% of the populace will be in the middle of the mean and in addition to 2.5 sigma and short 2.5 sigma and 99.7 % will be in the middle of in addition to and less 3 times sigma. These can be utilised as rule of thumb for computations of a variable that is normally distributed.

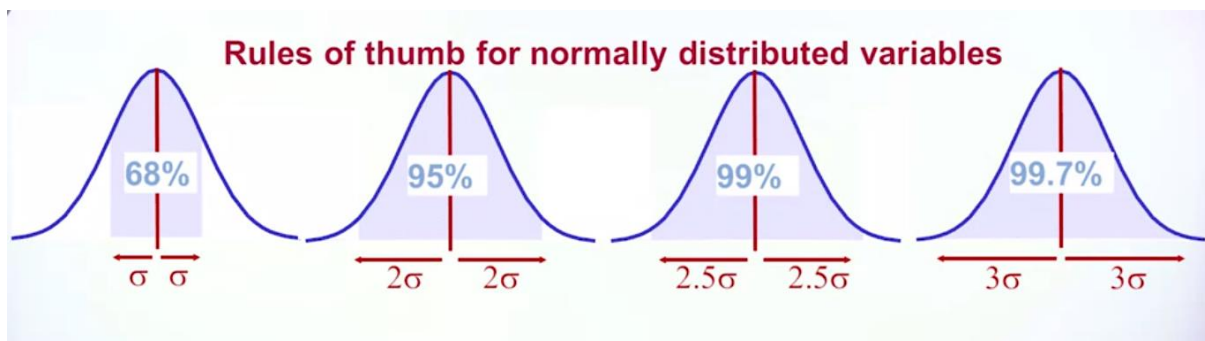


Figure 24: Rules of thumb – Normal distribution

The Weibull distribution is called a skewed conveyance. The Weibull distribution like in figure 24 is called skewed to one side, on the grounds that the tail is to one side. The Weibull distribution is regularly utilised for information, like, throughput times, and preparing times, in light of the fact that these are frequently skewed factors.

Numerous values are moderately little, and a few values are extremely high.

Probability Distributions – *Weibull Distribution*

Distribution Plot



Skewed distribution

Figure 25: Weibull distribution

The Lognormal distribution is normally distributed however is positive and skewed. It is a continuous probability distribution of an arbitrary variable whose logarithm is normally distributed.

Probability Distributions – *Lognormal Distribution*

Distribution Plot

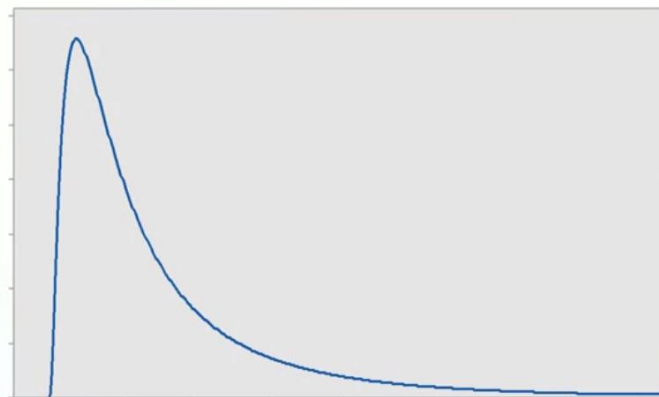


Figure 26: Lognormal distribution

XV. **PROBABILITY PLOTS**

The apparatus used to locate the best fitting distribution is known as the probability plot. This is possible utilising Minitab.

Figure 27 shows the probability plot. To check if the appropriation fits well, one needs to take a gander at the estimations, if the dots are on a straight line and however much as could

reasonably be expected between the two outside lines. In the figure we see that the normal distribution is a poor fit for the total handling time. Henceforth, the information's contrasted with the Weibull distribution. The data is not in between the lines.

Consequently, it does not fit well. So, take a glance at the Lognormal distribution. Here we see that the data lies in the middle of the red lines, so all the data is lognormally disseminated.

On the off chance that we look at all three charts, we can presume that a lognormal appropriation fits best to your information of THT.

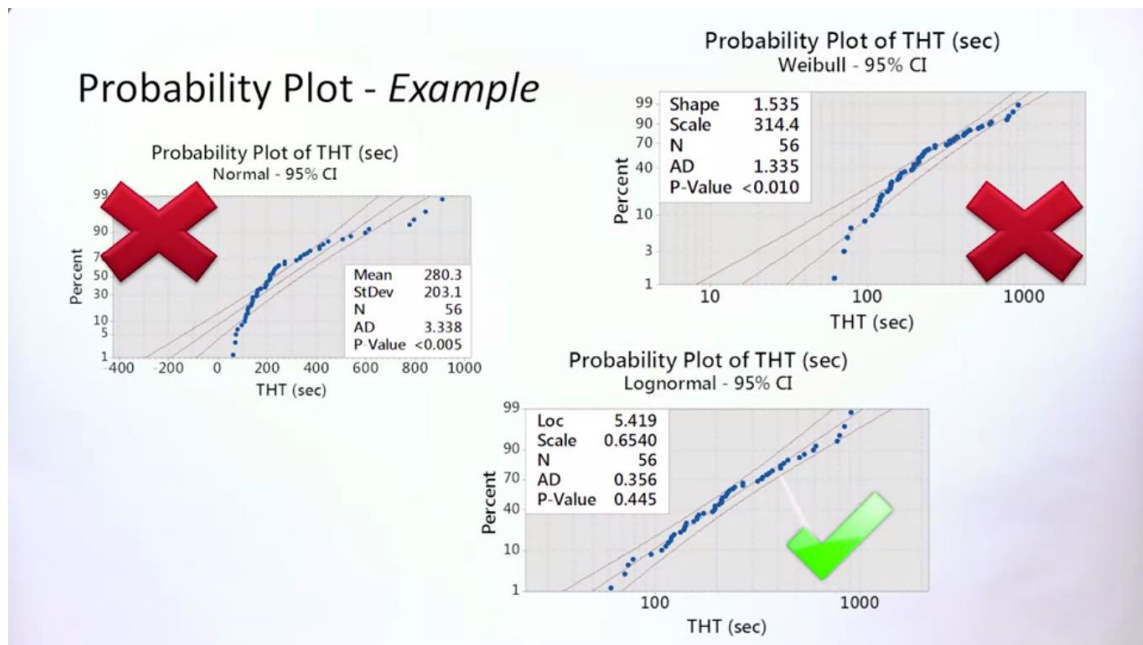


Figure 27: Probability plot

As a subsequent check, one can look at the AD esteems, this is the Anderson Darling measurement. It quantifies the separation among information and the hypothetical circulation, for example, the normal, Weibull, or lognormal dissemination. The lower the value, the better the fit. So, select the conveyance with the most reduced AD value.

In figure 27, the lognormal conveyance has a most minimal AD value and subsequently fits the information best.

XVI. EMPIRICAL CDF

To compute probabilities and rates, an instrument called the exact or empirical CDF is utilised.

In sigma venture, this device is utilised in the analyse stage. The empirical CDF is utilised to decide what number of rates of the cases meet help level understanding, or details?

Processing exact CDF utilising Minitab: In figure 28, delineating the empirical CDF of the total handling time, the blue bar speaks to the data of the example. What is more, the red curve speaks to the probability distribution which is a model for the populace. The red and the blue line covers for the lognormal distribution, consequently, is taken as a fit. The lognormal experimental CDF is determined for the data. Snap on the chart and snap on your

right mouse fasten and go to line of sight. This gives the cross to explore chart and thinking about what number of calls can be dealt with inside an objective of 240 seconds. Go to 240 seconds and afterwards go to the red line to anticipate the sum. So we see on upper left, we get the specific organise in 240 seconds, 54% of the calls are dealt with.

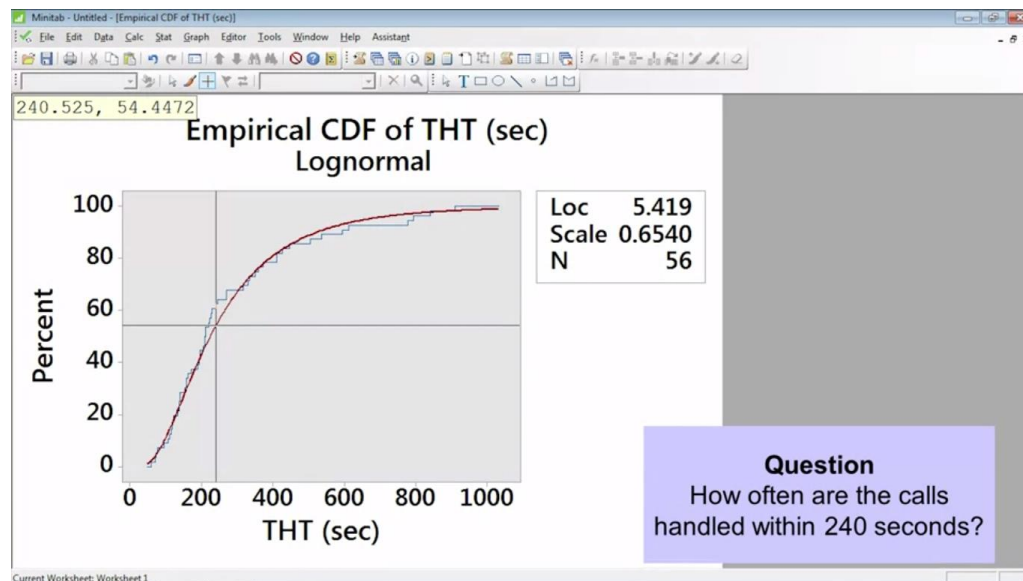


Figure 28: Empirical CDF example

XVII. DATA BASED TESTING IN LEAN SIX SIGMA

Information based testing of thoughts and improvements is one of the most significant standards of Lean Six Sigma. It is alluded to as proof based improvement activities. In DMAIC venture, relating the CTQ to the impact factor will be a piece of the improve stage where one creates and actualises successful improvement activities.

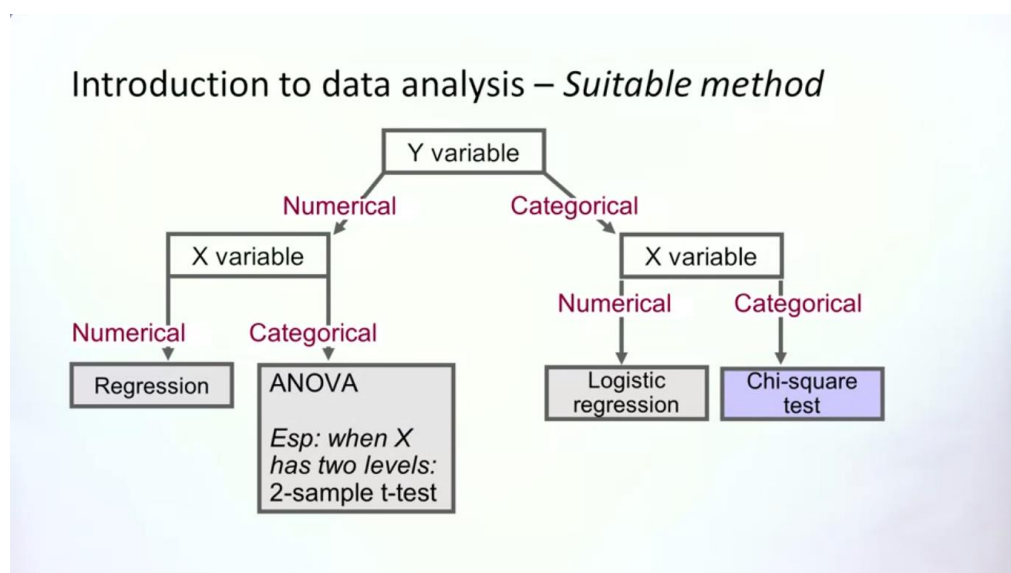


Figure 29: Data analysis- suitable method

XVIII. HYPOTHESIS TESTING

An estimate is just distinct. Subsequently, one regularly changes from estimation to testing. Testing implies that one claims something. What is more, information is utilised to test this claim. This claim is known as a hypothesis testing. For this one ascertains the p-value. The P-value is a probability, that is the reason it is called p-value. This p-value will permit choosing whether to reject null hypothesis.

On the off chance that the p-value is small enough, this implies one has discovered a huge impact, that H_0 is impossible and that you reject the null hypothesis. If a p-value is sufficiently little, lesser than 0.05 stating that we reject the null hypothesis.

On the off chance that the p-esteem is bigger than this edge of 0.05, we do not reject the null hypothesis.

XIX. CASUALITY

It is about circumstances and logical results connections. This implies when you change your X variable and nothing else your Y variable will change, therefore. Be that as it may, this is not the situation consistently. You can expect a causal relationship off course. You can relate two factors to one another that are random however change a similar route after some time. You can have a third factor that makes two factors change simultaneously. Or then again you can have a fundamental variable which is really the reason for the adjustments in Y rather than your underlying X variable.

You can abstain from making these blunders by playing out a controlled test or utilizing extra writing that discloses to you the circumstances and logical results relationship. Coherent reasoning may likewise assist with abstaining from making these mistakes.

XX. ANOVA ANALYSIS

ANOVA represents analysis of variance. ANOVA is the suitable measurable method when the Y variable is numerical, and the X variable is categorical. ANOVA can be acted in three stages; the initial step is to sort out information. The X and Y factors ought to have their different column. Each row in data set needs to contain one unit. With information organised, one can go to stage two and really play out the ANOVA. The third step consists of a presumption check on the grounds that after playing out an ANOVA, one needs to confirm the reliability of the outcomes.

ANOVA is a factual strategy so you can discover it under the menu Stat. Under ANOVA take the One-Way option. Minitab asks, what is Response? That is the Y variable or the CTQ? Presently the following inquiry is what's Factor? All things considered, factor is impact factor or X variable. Go to Graph and request an Individual value plot. Uncheck different plots. Under Options, uncheck the Assume Equal Variance option.

ANOVA – Hypothesis test

Analysis procedures often result in p-values. **P < 0.05**

That is because they are *hypothesis tests*.

Null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ❌

Alternative hypothesis $H_a: \mu_i \neq \mu_j$ for some i and j ✅

Figure 30: ANOVA – Hypothesis testing

The R-squared consistently lies somewhere in the range of zero and 100% and tells the amount of the variation in Y variable is clarified by X variable. The R squared estimates the effect of the impact factor on the reliant variable. If it is bigger, we can say the impact factor is progressively significant or pertinent, it is a major fish. In the event that the R squared is little this implies the informative intensity of the impact factor is feeble and that imperative impact factors are absent at the end of the day; this impact factor is a little fish.

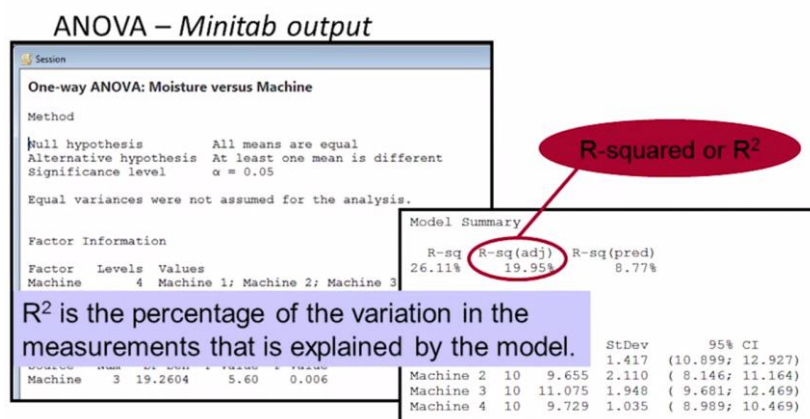


Figure31: ANOVA – R squared

The residual analysis is an imperative piece of any factual strategy. To approve the presumptions of the ANOVA investigation, we will check if the residuals are typically dispersed and if there are any anomalies or different inconsistencies present. In figure 31, Minitab gives the yield for residual analysis and it shows the data is normally dispersed.

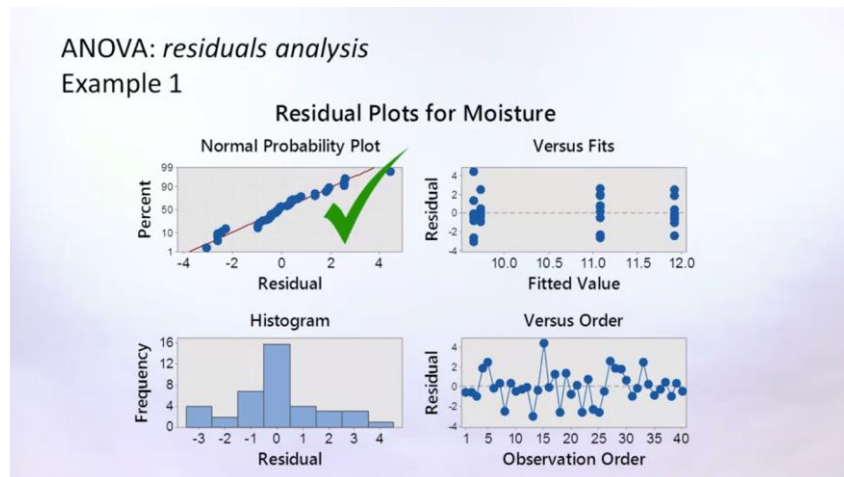


Figure 32: ANOVA- Residual analysis

To check for abnormalities, in figure 33 we see there are no exceptions or bizarre examples present. This implies this supposition that is likewise fulfilled, and that the first examination is legitimate.

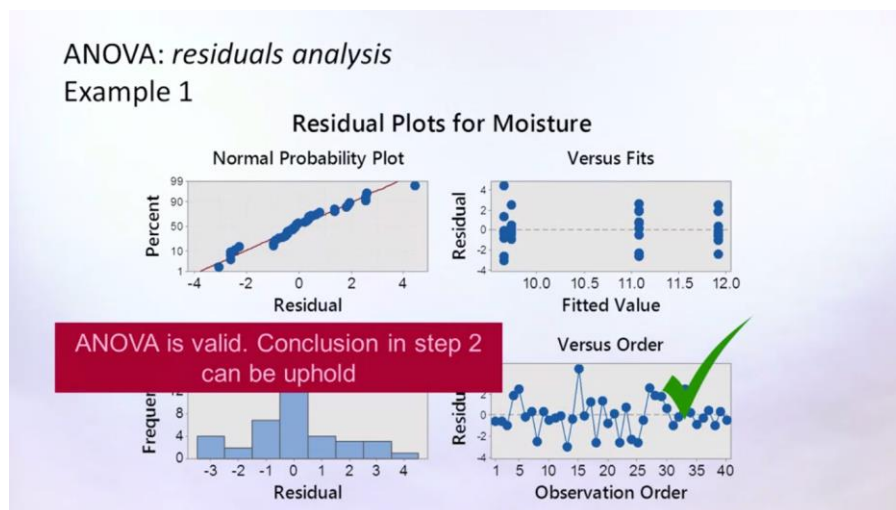


Figure 33: ANOVA- Residual analysis

In figure 34, we find in the probability plot that the residuals are not typically appropriated. Furthermore, in the line plot, we see exceptions in the residuals. This implies if these were residuals, the presumptions of the ANOVA are violated. This infers the end in step two would not have been legitimate, or if nothing else they are not exact. If so, play out a Kruskal-Wallis examination.

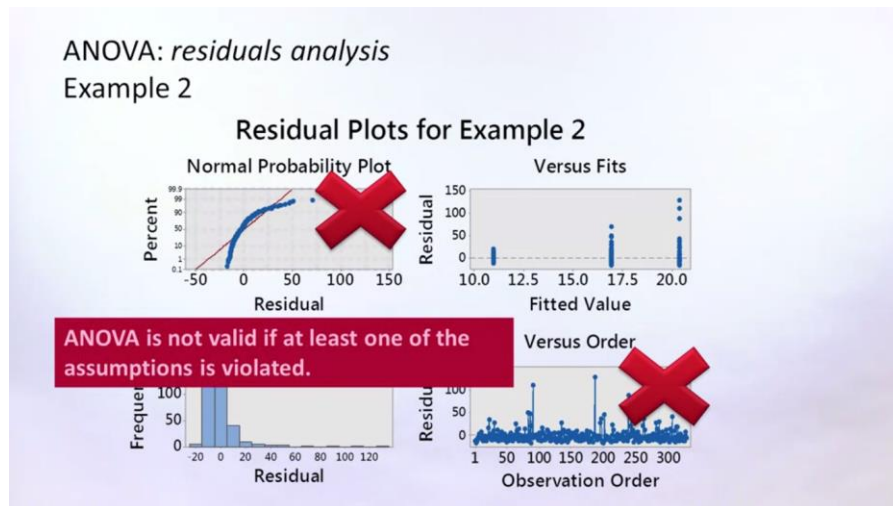


Figure 34: ANOVA- Residual analysis

XXI. KRUSKAL- WALLIS ANALYSIS

The Kruskal-Wallis test is a non-parametric test. This implies there is no distribution accepted on the residuals. This test ought to be performed when the basic suppositions of the ANOVA examination are not met. To decipher the yield, take a gander at the medians for each gathering and obviously, along with the p- value to check whether this distinction is statistically significant.

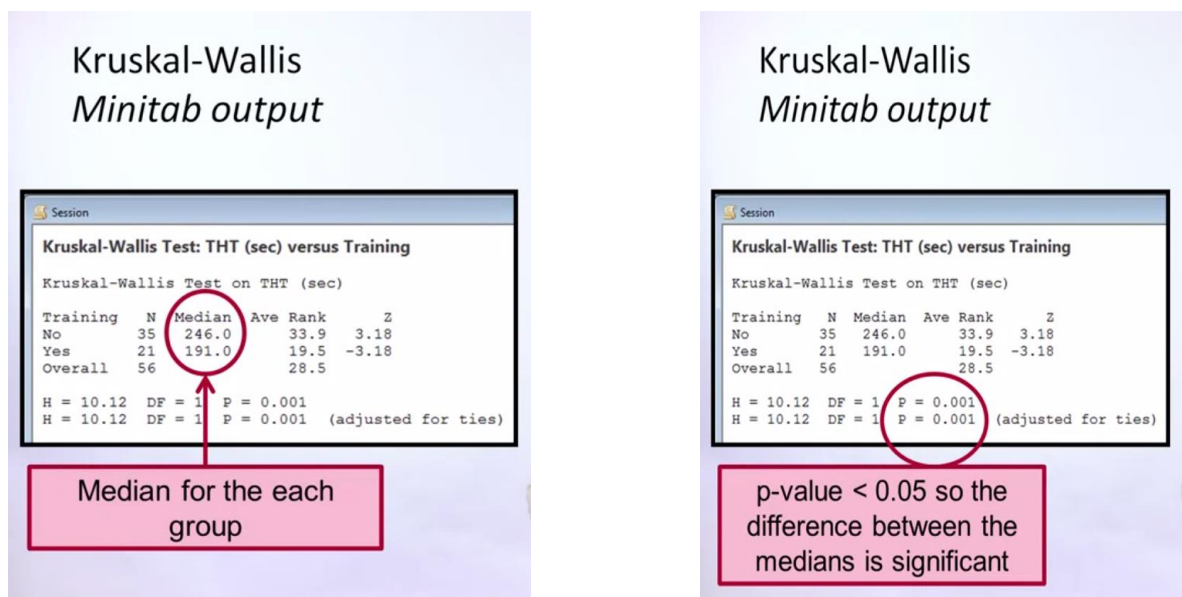


Figure 35: Kruskal- Wallis analysis

XXII. TWO SAMPLE t-TEST

The two-sample t-test is a technique to look at methods for two gatherings. Under Basic Statistics, one will locate the two-Sample t-test. In figure 36, the assessed distinction between

the mean of compost An and B is 2.74 and the p-value is 0.015. The p-value is lesser than 0.05, henceforth one can infer that the manure significantly affects the yield.

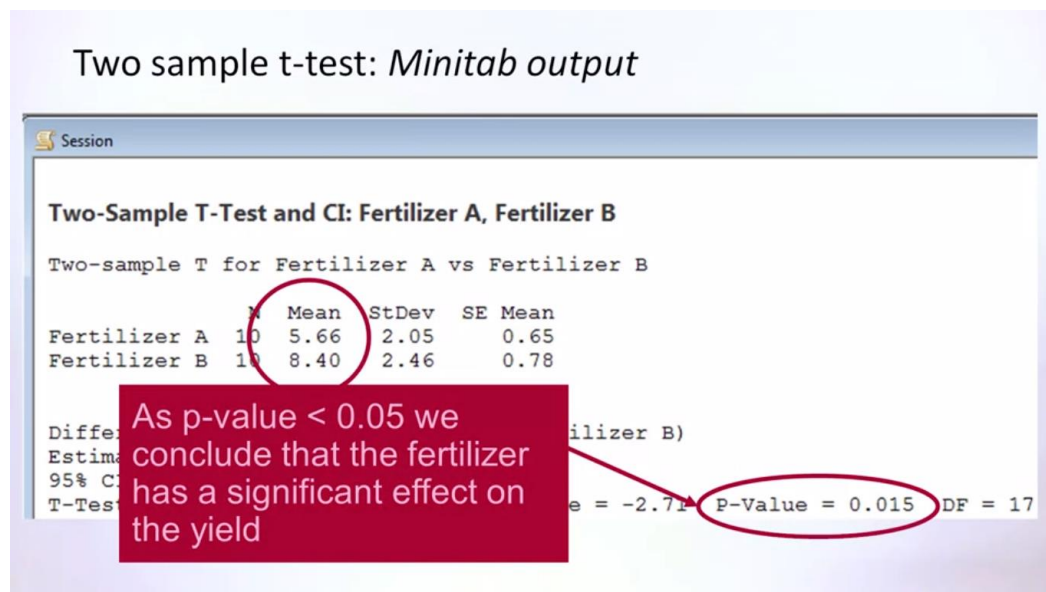


Figure 36: Two sample t- test

XXIII. TEST FOR EQUALITY OF VARIANCES

Expanding variability consistently debases the exhibition of a performance framework. This is the law of variation. This law, as talked about by Hopp and Spearman, shows that understanding variety is significant for the nature of a procedure. For testing of equivalent changes, go to the details menu since this is statistical analysis. On the detail menu, go to ANOVA. What is more, there, one can discover the test for equivalent differences.

In figure 37, the appraisals for the standard deviation in dampness rate for the machines are given. To check if the contrasts between these are measurably critical, take a gander at the p-value, figure it as per Levene's strategy. For this situation, the p-value is generally large. It gives us that there is a 32.5 rate chance that the distinction between the machines is an opportunity variance. This implies there's no proof that there is a distinction in variety among machines and we can't make any determinations on what the machine delivers most reliably.

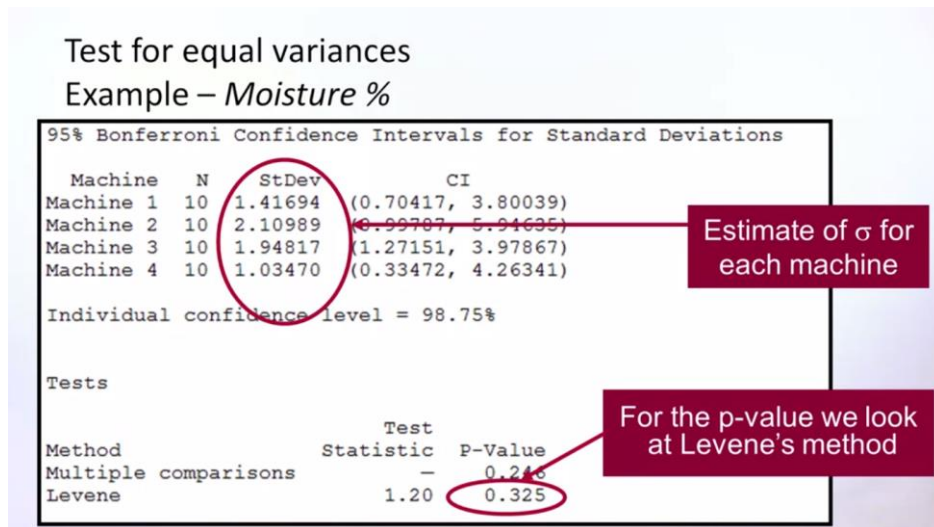


Figure37: Test for equality of variance

XXIV. CORRELATION

Correlation is a list that communicates how solid a connection between two numerical variables is. This list is consistently between plus 1 and minus 1.

In figure 38, A and B show a positive connection. This implies higher value of one of the variables will for the most part relate to higher values of other variable. In figure 39, A and B show a negative relationship which implies that the variables are corresponded the other way. So higher values of one of the factors for the most part relates to bring down estimations of the other variable.

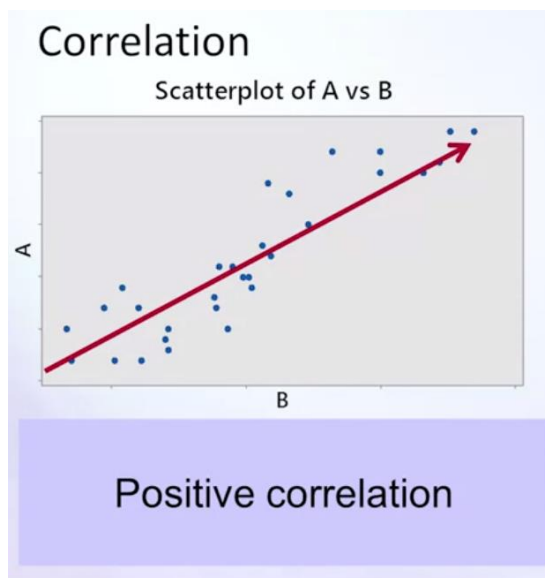


Figure 38: Positive correlation

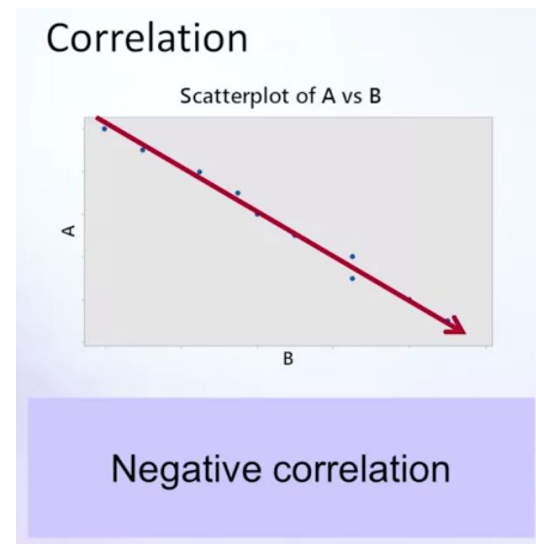


Figure 39: Negative correlation

In figure 40, the connection which equivalents to 0.93 delineates a solid, positive relationship, as the line is upward inclining, the relationship is certain. What's more, in light of the fact that the dots are exceptionally near the line, the connection is likewise near 1. The relationship is sure, yet it is moderately powerless in light of the fact that the spots are farther away from the line. The connection is 0.31, which we call a feeble positive relationship.

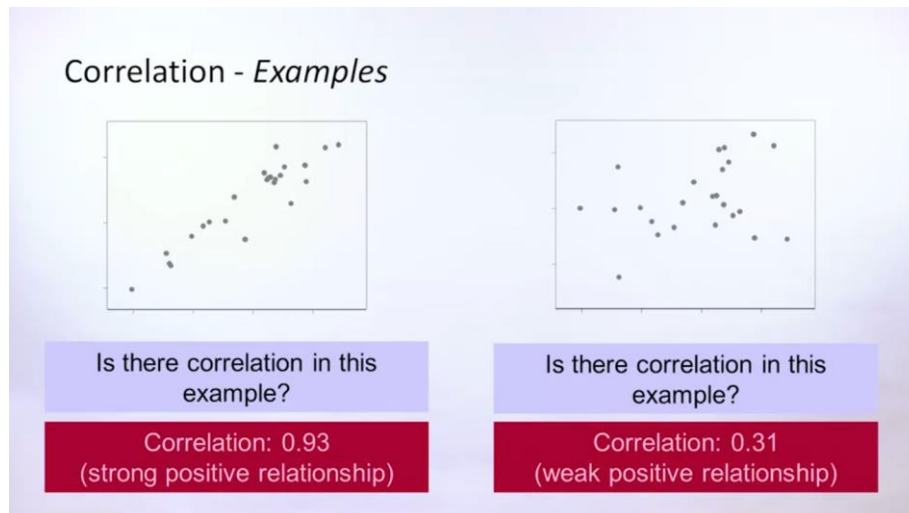


Figure 40: Strong/Weak positive correlation

Figure 41 shows curved relationship, an alleged non-linear relationship. In this way, we state that there is no relationship since connection is not a straight measure.

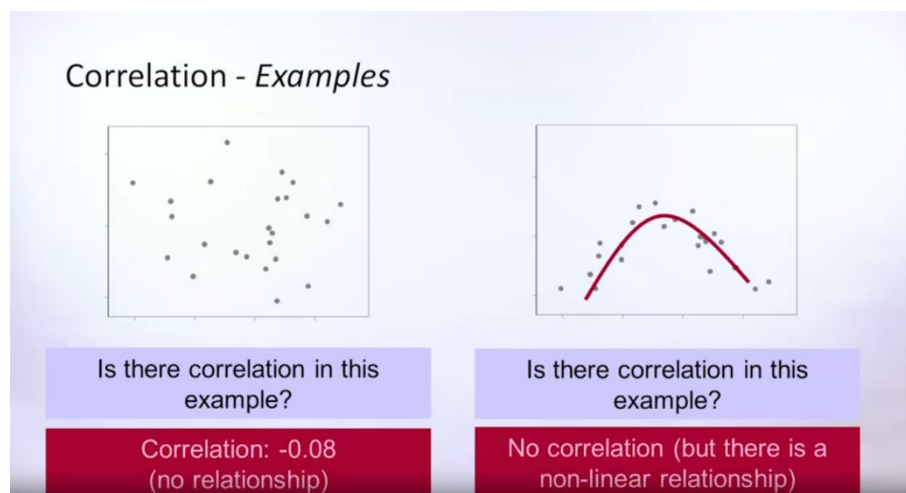


Figure 41: No correlation

In this manner relationship coefficient from 0 to 0.4 implies that there is no or a frail connection between the factors. A connection coefficient from 0.4 to 0.8 implies that there is a moderate relationship. A coefficient from 0.8 to 1 implies that there is an exceptionally solid relationship.

If the coefficient is 1, we talk about an ideal relationship between the two factors. In any case, as a rule, it seldom happens that two factors relate consummately.

XXV. **REGRESSION ANALYSIS**

Regression analysis is tied in with finding the best fitting straight line for two numerical factors in the data. A line can be portrayed scientifically by the recipe $y = a + bX$. Regression analysis implies finding the A and the B that gives the best fitting line. Regression analysis comprises of four stages. The initial step is making a fitted line plot to picture information. The subsequent advance is to play out the principle regression analysis.

At this progression, look whether the relationship is factually huge, and assuming this is the case, how solid the relationship is. The third step, you perform residual analysis. The fourth step is development of prediction interval.

In figure 42, we see the fitted line plot. We additionally observe the equation for the fitted line which is equivalent to the capacity $Y = a + bX$. The 'a' speaks to the constant of the line, which is a value when X equivalents to 0. Along these lines, with 0 stops, we will in any case anticipate 6.3 deformities. The 'b' speaks to the slant of the line, this is the steepness of the line. For this situation, if the quantity of stops is expanded by 1, the quantity of broken tea sacks is expanded by 1.944. Nonetheless, we despite everything need to decide whether these discoveries are because of possibility or on the off chance that they are auxiliary impact.

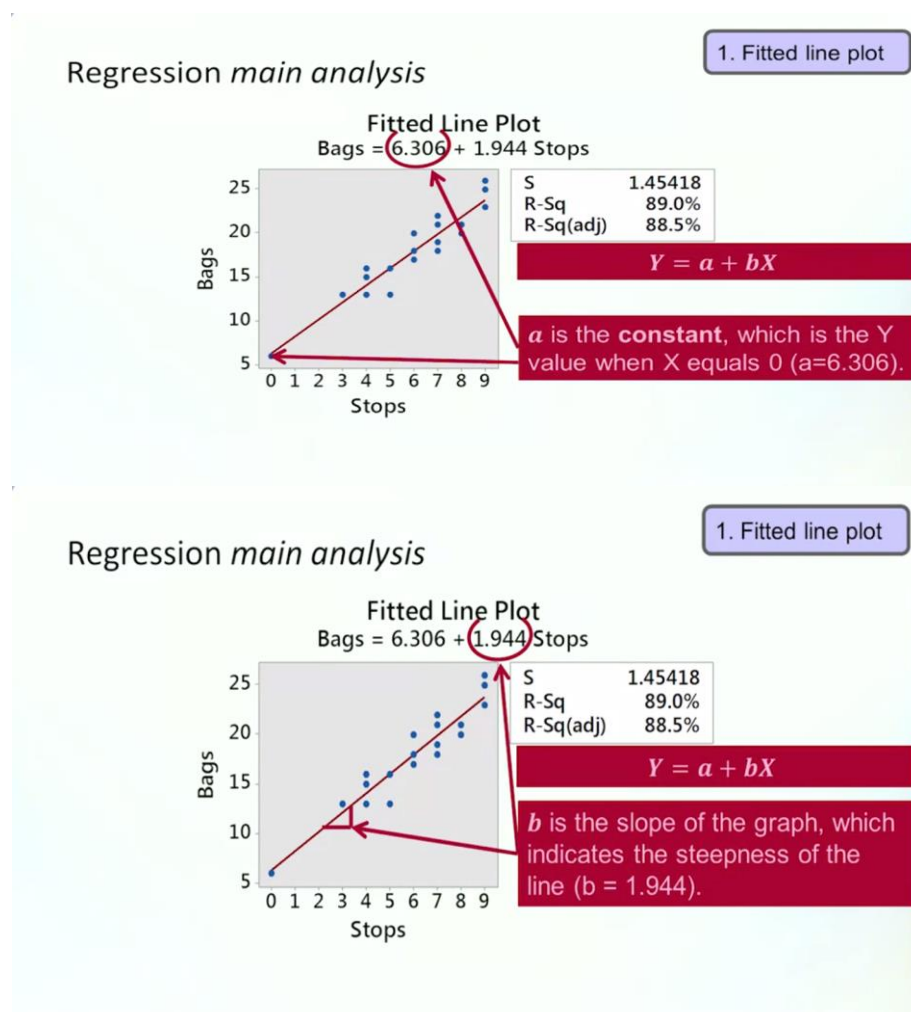


Figure 42: Regression analysis

For this, we need to play out the fundamental piece of the regression analysis. This is stage two, figure 43 portrays the yield in a meeting window of Minitab. We see that the P value is equivalent to 0 and it is littler than 0.05. Subsequently, we infer that there is a critical connection between stops and sacks. P-values are a piece of hypothesis tests. The null hypothesis expresses that there is no huge connection between the two factors. The alternative hypothesis expresses that there is a huge relationship. In the event that the p-value is lesser than 0.05, the null hypothesis can be dismissed, and the alternative hypothesis can be upheld.

Since we found an exceptionally little p-value, the alternative hypothesis can be supported. If the p-value would be bigger than 0.05, it implies that the outcomes are either because of possibility or that we did not accumulate enough information to demonstrate them.

Our R squared is 89%, which is very enormous. This implies the quantity of creation stops is a solid indicator for various broken tea sacks. The impact factor stops is a major fish. On the off chance that your R squared would have been little, it implies that there is a great deal of extra variety in your Y variable that remaining parts unexplained. The impact factor will be a little fish.

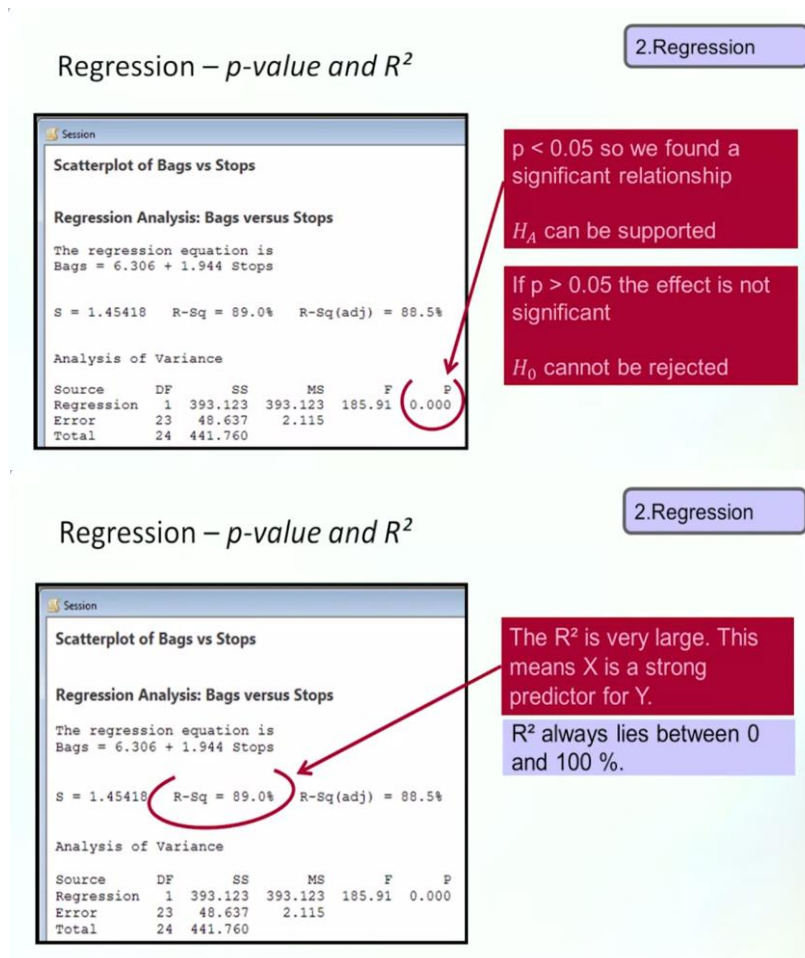


Figure 43: p -value and R squared

In figure 44, the upper left chart, we have a p -value that is low, and a high R squared. The impact vector has a critical impact and is a major fish. In the upper right diagram, the p -value is still low. Nonetheless, the R square is lesser since the information is more broadly spread around the line. Along these lines, more variety in the Y variable stays unexplained and X is a smaller fish. In the base left model, we see an extremely high p -value. There is no verification of a connection between the factors, henceforth R squared has no understanding. The last model shows a p -value which is high. So, there is no verification of our connections between the factors.

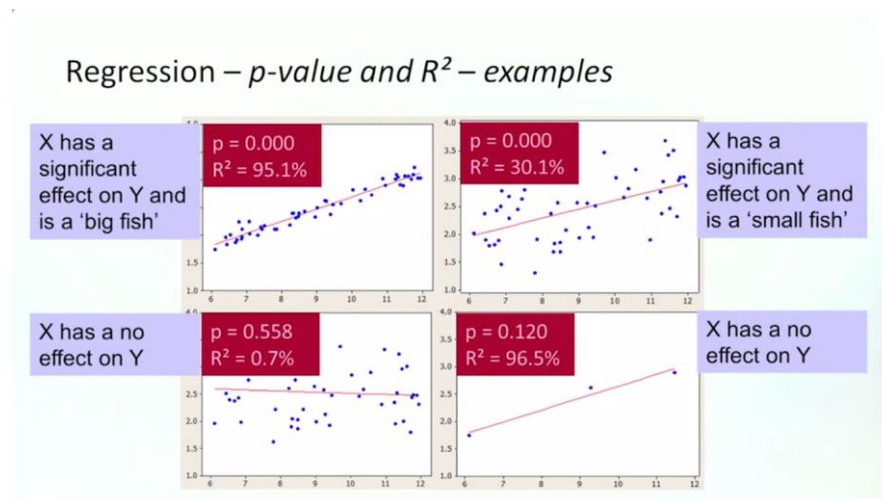


Figure 44: p -value and R squared – examples

Regression residual analysis: After the regression analysis, we need to check the basic suppositions of regression analysis. What is more, this is known as the residual analysis. We need to check if the residuals are normally distributed and if the residuals do not show any exceptions or different inconsistencies.

In figure 45, there is a distinction between the data and the fitted worth, and this distinction is known as the residual. Residuals are determined by taking away the observed value, the information, from the fitted worth, the estimated residual line. To play out a residual analysis, one can go to Minitab. For that, return to the relapse menu, which is found under detail, relapse, and the fitted line plots.

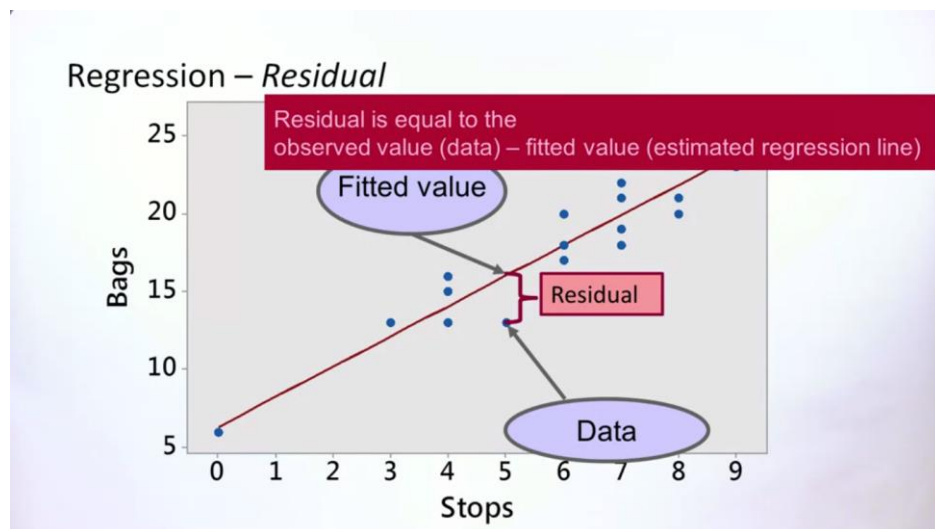


Figure 45: Regression – residual

In figure 46, we must check whether the residuals are normally circulated? They are. Presently, move onto the subsequent step. Do you see any outliers? It does not look like there are any. It implies that we can believe the ends from our regression and examination as the suspicions are fulfilled.

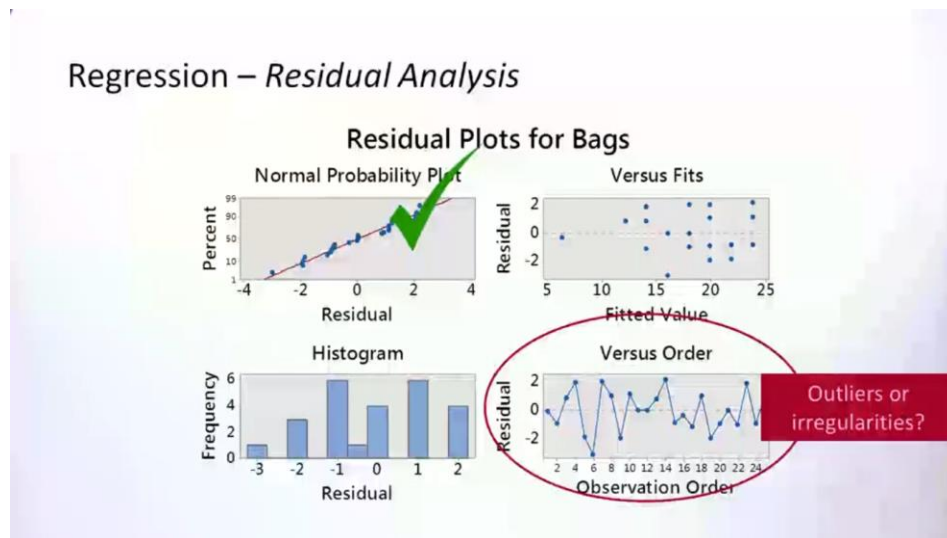


Figure 46: Regression- Residual analysis

Regression prediction interval: The prediction value can be utilised to decide the degree of impact factor, which guarantees a certain CTQ execution. This should be possible in Minitab by the utilisation of cross hairs, or the equation, to decide the necessary estimation of the impact factor

To ascertain the prediction interval, return to Regression examination menu, which is found under Stat. Under Option is the place you will discover the prediction interval. Snap it to get prediction interval and note that Minitab consequently works with a 95% level. Figure 47 portrays the will be the computed prediction interval. The prediction interval is the fitted line give or take multiple times the standard deviation of the residuals. So, 95% of the estimations ought to be inside this interval. Also, more than two level of the estimations will be over the stretch and obviously, more than two rates will be beneath the span. Likewise, if you rehash the estimations as often as possible, 97.5% of the estimations will be underneath the upper line of the stretch. In this way, if we need the limit of broken tea Bags to be 15 in 97.5% of the days, we must have all things considered 3 Stops for each day. This is processed by taking your mouse over the focal diagramming territory and going to Crosshairs. This will give you the kind of cross to explore the chart, and your directions of the focal point of the cross are given on upper left of the diagram

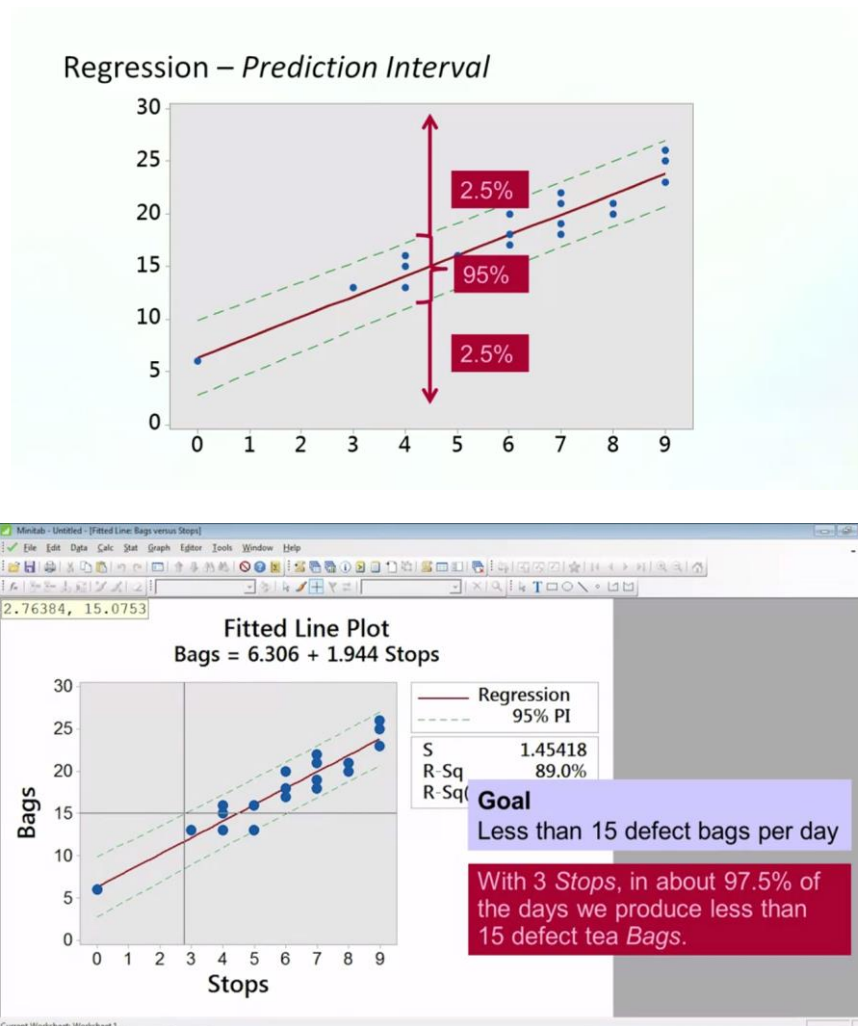


Figure 47: Prediction interval

XXVI. QUADRATIC REGRESSION

If the linear regression analysis doesn't meet the presumption, you should apply an alternative procedure. One of these options is the quadratic regression analysis. This procedure fits a curved line to your data rather than a straight line. After you have played out the regression analysis, you generally need to play out the residual analysis again to check whether the suppositions are met.

XXVII. CHI- SQAURE ANALYSIS

If you have a Y variable is categorical and a X variable or the influence factor that is numerical, we need to play out a chi-square investigation. How about we make a glance at the strides of the chi-square investigation. The initial step is to make a cross tabulations. You can either do this by duplicating a cross table into worksheet, or by having it done by Minitab. The subsequent advance is to investigate the p-value.

Indeed, the cross-tabulation yield in figure 48 gives us a great deal of data. The top lines give us what number of these patients were treated by every authority at division A1. For instance, at division A1, 52 patients were treated by Specialist Q.

Chi-square analysis - Output

Count	Rows: Department Columns: Specialist			
Total amount of patients treated by each specialist at department A1	P	Q	R	All
	81	52	78	211
	38.39	24.64	36.97	100.00
	82.65	77.61	75.00	78.44
	17	15	26	58
	29.31	25.86	44.83	100.00
	17.35	22.39	25.00	21.56
	98	67	104	269
	36.43	24.91	38.66	100.00
	100.00	100.00	100.00	100.00

Cell Contents: Count
% of Row
% of Column

Figure 48: Chi-Square output- Count

The line beneath in figure 49 shows us the level of patients in office A1 that were treated by one of the specialists. For instance, 38% of the patients that were treated in office A1 was treated by specialist P.

Chi-square analysis - Output

Row percentage	Rows: Department Columns: Specialist			
Percentage of patients that were treated by each specialist on department A1	P	Q	R	All
	81	52	78	211
	38.39	24.64	36.97	100.00
	82.65	77.61	75.00	78.44
	17	15	26	58
	29.31	25.86	44.83	100.00
	17.35	22.39	25.00	21.56
	98	67	104	269
	36.43	24.91	38.66	100.00
	100.00	100.00	100.00	100.00

Cell Contents: Count
% of Row
% of Column

Figure 49: Chi- Square output – Row percentage

The base column in figure 50 shows us the level of patients that were treated by specialists on one of the offices. So, for instance, 82% of the considerable number of patients were treated by specialist P at division A1.

Chi-square analysis - Output

Column percentage				
Percentage of patients that were treated on department A1, out of all the patients treated by each specialist				
Rows: Department Columns: Specialist				
	P	Q	R	All
A1	81	52	78	211
	38.39	24.64	36.97	100.00
	82.65	77.61	75.00	78.44
A4	17	15	26	58
	29.31	25.86	44.83	100.00
	17.35	22.39	25.00	21.56
All	98	67	104	269
	36.43	24.91	38.66	100.00
	100.00	100.00	100.00	100.00
Cell Contents: Count				
% of Row				
% of Column				

Figure 50: Chi-Square output – Column percentage

In the wake of deciphering the cross tabulation, the subsequent advance is to interpret the p-value. In figure 51, the p-value is more noteworthy than 0.05, subsequently null hypothesis cannot be dismissed

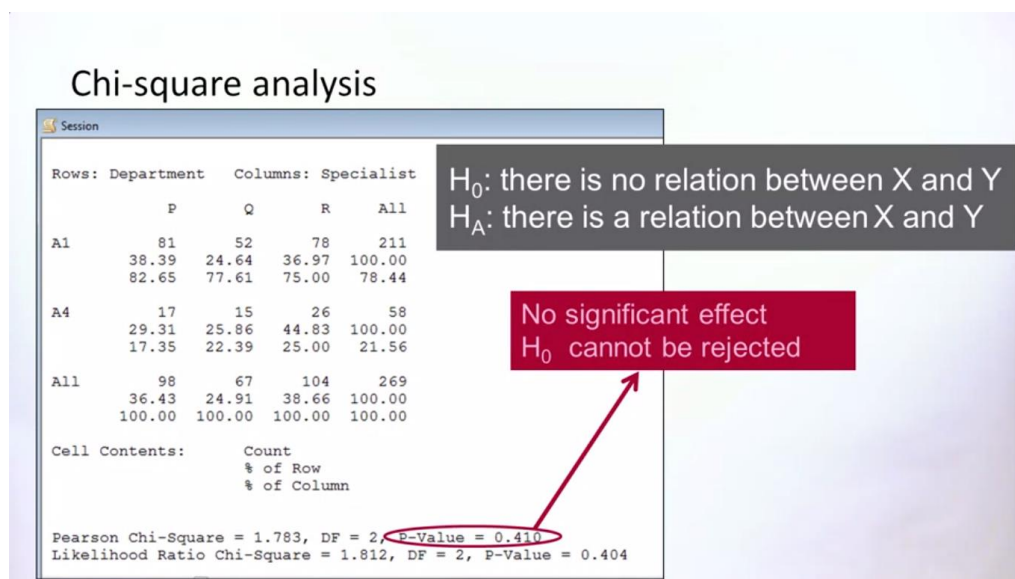


Figure 51: Chi-Square p-value

XXVIII. LOGISTIC REGRESSION

At the point when you have a categorical Y variable and a numerical X variable, the logistic regression is calculated. Logistic Regression comprises of three stages. In the initial step you arrange your information in the event trail manner. In the second step you make a fitted line plot. What is more, in the third step, you check if the outcomes are huge and can be summed up to the populace.

The event trial resembles the one in figure 52. In the main segment you put every one of your qualities for your X variable, which is the entire time estimated in minutes. In the third segment you note the trials, which is the quantity of calls that got a normal hold time for the quantity of minutes in the primary section. The subsequent section is then occasions segment which takes note of the quantity of break of the all-out number of times the event occurred. For this situation, an occasion is that someone has hung up. So, it is the quantity of individuals that with this specific holding up time have hung up.

Logistic Regression - Example

Event/trial format		
HoldTime	HungUp	Calls
1	1	19
2	0	15
3	2	12
4	1	15
5	5	14
6	2	13
7	8	17
8	10	21
9	9	18
10	3	13
11	7	13
12	10	16
13	7	15
14	9	19
15	18	22
16	19	25
17	12	17
18	13	16
19	21	21
20	14	15

Example
data_files_DA-LSS.xls
[Callcenter](#)

X variable
All possible outcomes of the X variables

Trials
Number of times that this level of the X variable occurred

Events
Number of times that the event occurred for this level of X.

Figure 52: Event trial format

In the wake of changing over the data into event trial format, play out the logistic regression. To play out a logistic regression, we, go to the Stat menu and under Regression you will locate the Binary Fitted Line Plot. For stage 2, we need to investigate the fitted line plot which resembles the one in figure 53. We see that the possibility of hanging up increments when the hold time increments. At the point when the normal hold time is around five minutes the likelihood of a hanging up is 0.2, which implies that 20% of the individuals that hear that they have hold as long as five minutes will hang up. We additionally observe that many did not fit a straight line yet that it is a marginally twisted line which is kind of a S bend. This is on the grounds that the likelihood of hanging up can never have an incentive underneath zero, or over one, as it is a likelihood.

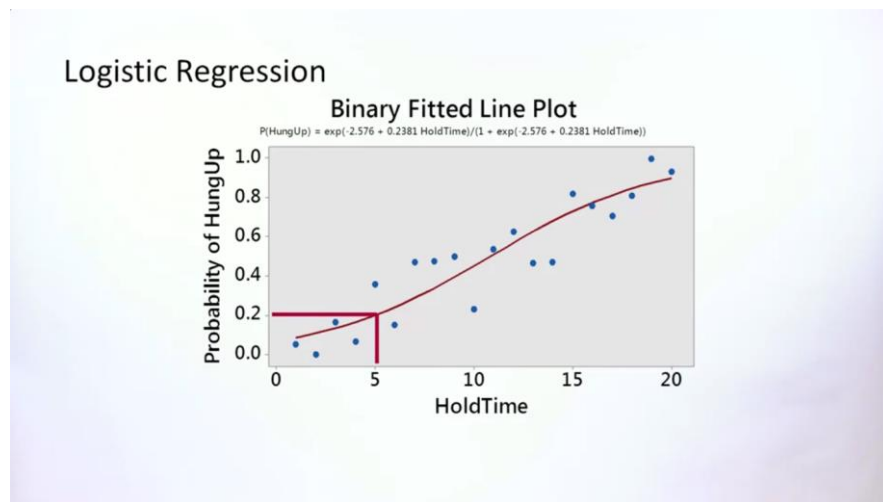


Figure 53: Binary fitted line plot

Presently, we need to check if the results are significant. This is stage three. To do this, we need to investigate Menu Top's meeting window. The output in a meeting window gives the summary of the analysis. In figure 54, the p-value demonstrates that the hold time is a huge impact factor on your hanging up or not.

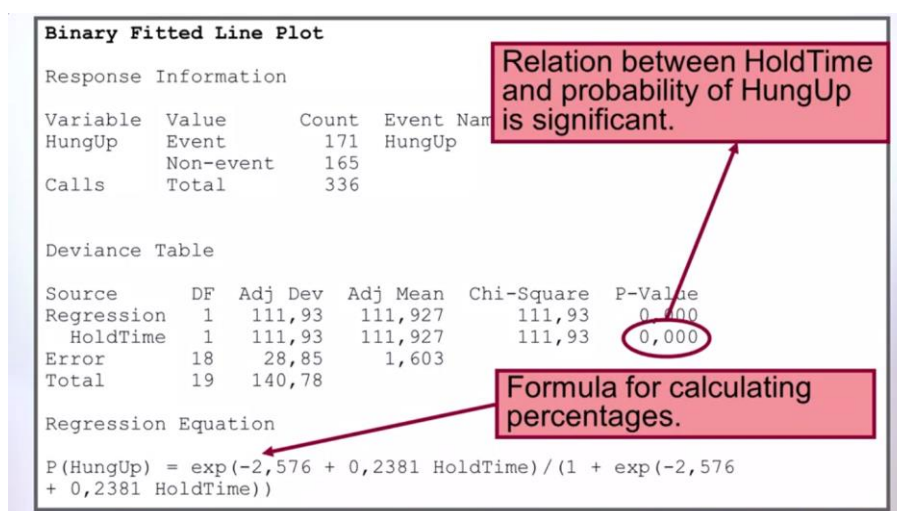


Figure 54: Binary fitted plot – p- value

CONCLUSION

Toward the finish of the course, I briefly know about Lean Six Sigma and how data analytics methods are valuable inside Lean Six Sigma improvement process. I am ready to analyse and interpret information accumulated for any improvement process needed in an organisation and have built up the effectiveness to utilize Minitab to break down the information, with different information expository devices and interpret the result.

REFERENCES

1. Course content- Coursera/ Data Analytics In Lean Six Sigma
2. The figures have been developed by IBIS UvA, the Lean Six Sigma institute of the University of Amsterdam (www.ibisuva.nl)