

A Comprehensive Guide to Advanced Business Statistics

Fairness

Team Spirit

A Comprehensive Guide to Advanced Business Statistics



Website: www.skyfox.co

Email: skyfox@skyfox.org.in



A Comprehensive Guide to Advanced Business Statistics



Dr.R.S.Thangeswari



A Comprehensive Guide to Advanced Business Statistics

AUTHOR INFORMATION:

Dr.R.S.Thangeswari, M.Com, M.Phil, B.Ed., Ph.D.,
Assistant Professor,
Department of Commerce,
Kamaraj College,
Thoothukudi- 628 003
Tamilnadu, India.

First Edition : August 2020
Cover Art and Design : Author
ISBN : 978-93-90357-28-4
DOI : <https://doi.org/10.22573/spg.020.BK/S/025>
Copyright : © 2020 by Authors

Creative Commons Attribution-Share Alike 4.0 International License

You are free: to Share — to copy, distribute and transmit the work; to make commercial use of the work under the following conditions:

Attribution — you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that, they endorse you or your use of the work).

No Derivative Works — you may not alter, transform, or build upon this work.

Publisher Contact &Support

Skyfox Publishing Group

#987, Medical College Road

Thanjavur-613004

Tamil Nadu, India.

Phone: +918300123232

Email: skyfoxpublishing@gmail.com / skyfox@skyfox.org.in

Website: www.skyfox.co

Headquarters &Marketing Office

Skyfox Publishing Group

333 Cedar Street, PO Box 208002,

New Haven, United States.

CT 06520-8002.

Tel: 203.737.5603 / Fax: 203.785.7446

Email: skyfoxpublishing@gmail.com / skyfox@skyfox.org.in

Website: www.skyfox.co

A COMPREHENSIVE GUIDE TO ADVANCED BUSINESS STATISTICS

Dr.R.S.Thangeswari

PREFACE

I have been educating statistics to postgraduate commerce and business students for almost 15 years. When I took the course as postgraduate, I realized that the learners are struggling to get the basic knowledge about statistics. Though they are studying post graduate, to grasp the edge of leaves cumbersome remains calm. Normally a prejudice among a new topics especially the subject like statistics attains the same upon the non-major students.

“If beginning crooked, ending crooked also” at the beginning even the veteran statistician face this difficulty how am I going to transform their entire negative preconception into a positive one. In a field of transforming learners through education, it’s a challenging task of all the teachers to make the learners to realize in due course how easy the subject is. And this guide assists those educators to achieve their challenging task in a successful manner.

“First impression is a best impression” Non major undergraduate commerce and business students who is selecting the electives supposed to study statistics in their post graduate without any choice. Scenario seems difficult while come to know about statistics subject. Here this guide waits to serve you to study statistics. Like the way my preface stands learners can able to comprehend advanced business statistics. In this comprehensive guide I have prepared a coherency of various statistical topics. As per students own perspective I have geared up this comprehensive guide to business statistic book.

- DR.R.S.Thangeswari, M.Com, M.Phil, B.Ed., Ph.D.,
Assistant Professor,
Department of Commerce,
Kamaraj College,
Thoothukudi,
Tamilnadu, India.

CONTENTS

Chapter – I	Theoretical Distribution
Chapter – II	ANOVA Table
Chapter – III	Chi – Square Test
Chapter – IV	Statistical Decision Theory

Chapter - I
Theoretical Distribution
Or
Expected (or) Mathematical Distribution

1. Binomial Distribution
2. Poisson Distribution
3. Normal Distribution

1. Binomial Distribution

P	=	Probability of success
q	=	Probability of failure
q	=	1 - P
n	=	no. of trials - items

- i) Mean of the Binomial Distribution

$$\bar{X} = np$$

- ii) Standard Deviation

$$\sigma = \sqrt{npq}$$

- iii) Movement co-efficient of skewness :

$$\beta_1 = \frac{(q - p)^2}{npq}$$

- iv) Co-efficient of kurtosis :

$$\beta_2 = 3 + \frac{1 - 6pq}{npq}$$

$$P(r) = nC_r \times q^{n-r} \times p^r$$

$$NP(r) = N \times nC_r \cdot q^{n-r} \cdot p^r$$

1. If the probability of defective bolt is 0.1, find the mean and standard deviation for the distribution of defective bolt, the total of 500.

Movement co-efficient of skewness and kurtosis of the distribution.

- i) The mean of binomial distribution

$$\bar{X} = np$$

$$= 500 \times 0.1 = 50$$

ii) Standard deviation

$$\begin{aligned}\sigma &= \sqrt{npq} \\ q &= 1 - p \\ &= \sqrt{500 \times 0.1 \times 0.9} \\ &= 1 - 0.1 = 0.9 \\ &= \sqrt{45} = 6.708\end{aligned}$$

iii) Movement co-efficient of skewness :

$$\begin{aligned}\beta_1 &= \frac{(q - p)^2}{npq} \\ &= \frac{(0.9 - 0.1)^2}{500 \times 0.1 \times 0.9} \\ &= \frac{(0.8)^2}{45} = \frac{0.64}{45} = 0.01\end{aligned}$$

iv) Co-efficient of skewness:

$$\begin{aligned}\beta_2 &= 3 + \frac{1 - 6pq}{npq} \\ &= 3 + \frac{1 - 6(0.1 \times 0.9)}{45} \\ &= 3 + \frac{1 - 6(0.09)}{45} \\ &= 3 + \frac{1 - 0.54}{45} \\ &= 3 + \frac{0.46}{45} = 3 + 0.01 = 3.01\end{aligned}$$

Role:

$\beta_2 = 3$ Cavesis called as mesokartitic

$\beta_2 > 3$ Thecave is called as leppokurtic.

$\beta_2 < 3$ The caves is called as platykurtic

Here β_2 is > 3 (3.01 is > 3). The curve is called as leppokurtic.

2. The mean of the binomial distribution is 20 and σ is 4.

Calculate the value of n, p, q .

i) Mean of binomial distribution

$$\bar{X} = np$$

$$20 = np$$

ii) Standard deviation

$$\sigma = \sqrt{npq}$$

$$\begin{aligned}
 4 &= \sqrt{npq} \\
 4^2 &= npq \\
 16 &= npq \\
 16 &= 20q \\
 q &= \frac{16}{20} \\
 &= \frac{4}{5} = 0.8
 \end{aligned}$$

$$\begin{aligned}
 P &= 1 - q \\
 &= 1 - \frac{4}{5} = \frac{1}{5} = 0.2 \\
 20 &= np \\
 20 &= n(0.2)
 \end{aligned}$$

$$\frac{20}{0.2} = n$$

$$n = 100$$

3. 5 dice of through together 96 dice, the number of time 4,5,6 was actually through the experiment is given below. Calculate the expected frequently.

No. of dice showing	0	1	2	3	4	5
Frequency	4	10	21	35	18	8

$$NP_{(r)} = N \cdot nC_r \cdot q^{n-r} \cdot p^r$$

$$N = \Sigma F$$

$$= 96.5C_0 \cdot \left(\frac{1}{2}\right)^{5-0} \cdot \left(\frac{1}{2}\right)^0 = 3$$

No. of dice showing	Frequency	Expected Frequency
0	4	3
1	10	15
2	21	30
3	35	30
4	18	15
5	8	3

$$\begin{aligned}
 NP_{(1)} &= 96.5C_1 \cdot \left(\frac{1}{2}\right)^{5-1} \cdot \left(\frac{1}{2}\right)^1 \\
 &= 96.5 (0.0625) \cdot (0.5) \\
 &= 15
 \end{aligned}$$

$$\begin{aligned}
 NP_{(2)} &= 96.5 C_2 \cdot \left(\frac{1}{2}\right)^{5-2} \cdot \left(\frac{1}{2}\right)^2 \\
 &= 96.10 \cdot (0.125) \cdot (0.25) \\
 &= 30 \\
 NP_{(3)} &= 96.5 C_3 \cdot \left(\frac{1}{2}\right)^{5-3} \cdot \left(\frac{1}{2}\right)^3 \\
 &= 96.10 (0.25) \cdot (0.125) \\
 &= 30 \\
 NP_{(4)} &= 96.5 C_4 \cdot \left(\frac{1}{2}\right)^{5-4} \cdot \left(\frac{1}{2}\right)^4 \\
 &= 96.5 (0.5) \cdot (0.0625) \\
 &= 15 \\
 NP_{(5)} &= 96.5 C_5 \cdot \left(\frac{1}{2}\right)^{5-5} \cdot \left(\frac{1}{2}\right)^5 \\
 &= 96.1 (1) \cdot (0.03125) \\
 &= 3
 \end{aligned}$$

4. Out of 100 family that they 5 children each how many would you expect to have i) 3 boys ii) 5 Girls iii) 2 (or) 3 boys (\pm). Assume equal probability of boys and girls.

$$\begin{aligned}
 N &= 100, n = 5, p = \frac{1}{2}, q = 1 - p = \frac{1}{2} \\
 \text{i) } NP_{(3)} &= 100.5 C_3 \cdot \left(\frac{1}{2}\right)^{5-3} \cdot \left(\frac{1}{2}\right)^3 \\
 &= 100.10 \cdot 0.25 \cdot 0.125 \\
 &= 31.25 \\
 \text{ii) } NP_{(0)} &= 100.5 C_0 \cdot \left(\frac{1}{2}\right)^{5-0} \cdot \left(\frac{1}{2}\right)^0 \\
 &= 100.1 \cdot (0.03125) \cdot 1 \\
 &= 3.125 \\
 \text{iii) } NP_{(2)} &= 100.5 C_2 \cdot \left(\frac{1}{2}\right)^{5-2} \cdot \left(\frac{1}{2}\right)^2 \\
 &= 100.10 \cdot (0.125) \cdot 0.25 \\
 &= 31.25 \\
 \text{iv) } NP_{(3)} &= 100.5 C_3 \cdot \left(\frac{1}{2}\right)^{5-3} \cdot \left(\frac{1}{2}\right)^3 \\
 &= 100.10 \cdot (0.25) \cdot (0.125) \\
 &= 31.25 \\
 NP_{(2)} + NP_{(3)} &= 31.25 + 31.25 \\
 &= 62.5
 \end{aligned}$$

5. If hens of a certain breeding lay eggs of on 4 days have week, find how many days in extension of 200 days of poultry keeper with 8 hens is the expected to received at least 6 eggs.

$$N = 200, n = 8 \text{ hens}, p = \frac{4}{7}, q = \frac{3}{7}$$

$$\begin{aligned} \text{i) } NP_{(6)} &= 200.8 C_6 \cdot \left(\frac{3}{7}\right)^{8-6} \cdot \left(\frac{4}{7}\right)^6 \\ &= 200.28. (0.429)^2. (0.57)^6 \\ &= 200.28.0.184. 0.0343 \\ &= 35.34 \end{aligned}$$

$$\begin{aligned} \text{ii) } NP_{(7)} &= 200.8 C_7 \cdot \left(\frac{3}{7}\right)^{8-7} \cdot \left(\frac{4}{7}\right)^7 \\ &= 200.8. (0.429)^1. (0.57)^7 \\ &= 200.8. (0.429). 0.0195 \\ &= 13.42 \end{aligned}$$

$$\begin{aligned} \text{iii) } NP_{(8)} &= 200.8 C_8 \cdot \left(\frac{3}{7}\right)^{8-8} \cdot \left(\frac{4}{7}\right)^8 \\ &= 200.1. 1. 1. (0.57)^8 \\ &= 200.1. 1 (0.0114) \\ &= 2.23 \end{aligned}$$

$$\begin{aligned} NP_{(6)} + NP_{(7)} + NP_{(8)} &= 35.34 + 13.42 + 2.23 \\ &= 50.99 \end{aligned}$$

6. A sample of 3 items selected at random from a box containing 12 items of which 3 are defective. Find the possible no. of defective combination of the said 3 selected item along with probability of defective combination.

During cold epidemic in a factory, the probability of the worker catching cold is $\frac{1}{5}$. Find the probability of at least 7 workers catching cold, if 10 workers are selected at random.

$$n = 3, \quad p = \frac{3}{12}, \quad q = 1 - p = \frac{9}{12}$$

$$n = 10, \quad p = \frac{1}{5}, \quad q = 1 - p = \frac{4}{5}$$

$$\text{i) } NP_{(r)} = n C_r \cdot q^{n-r} \cdot p^r$$

$$\begin{aligned} NP_{(0)} &= 3 C_0 \cdot \left(\frac{9}{12}\right)^{3-0} \cdot \left(\frac{3}{12}\right)^0 \\ &= 1. (0.75)^3. 1 \\ &= 1. (0.422). 1 \end{aligned}$$

$$NP_{(0)} = 0.422$$

$$\begin{aligned} NP_{(1)} &= 3 C_1 \cdot \left(\frac{9}{12}\right)^{3-1} \cdot \left(\frac{3}{12}\right)^1 \\ &= 3. (0.75)^2. (0.25) \\ &= 3. (0.563). (0.25) \end{aligned}$$

$$\begin{aligned}
&= 0.422 \\
NP_{(2)} &= 3 C_2 \cdot \left(\frac{9}{12}\right)^{3-2} \cdot \left(\frac{3}{12}\right)^2 \\
&= 3 \cdot (0.75) \cdot (0.25)^2 \\
&= 3 \cdot (0.75) \cdot (0.25)^2 \\
&= 3 \cdot (0.75) \cdot (0.625) \\
NP_{(2)} &= 0.141 \\
NP_{(3)} &= 3 C_3 \cdot \left(\frac{9}{12}\right)^{3-3} \cdot \left(\frac{3}{12}\right)^3 \\
&= 1.1 \cdot (0.25)^3 \\
&= 1.1 \cdot (0.0156) \\
NP_{(3)} &= 0.156 \\
NP_{(0)} + NP_{(1)} + NP_{(2)} + NP_{(3)} &= 0.422 + 0.422 + 0.14 + 0.0156 \\
&= 1.006 \\
\text{ii) } NP_{(7)} &= 10 C_7 \cdot \left(\frac{4}{5}\right)^{10-7} \cdot \left(\frac{1}{5}\right)^7 \\
&= 120 \cdot (0.8)^3 \cdot (0.2)^7 \\
&= 120 \cdot (0.512) \cdot (0.0000128) \\
NP_{(7)} &= 0.000786 \\
NP_{(8)} &= 10 C_8 \cdot \left(\frac{4}{5}\right)^{10-8} \cdot \left(\frac{1}{5}\right)^8 \\
&= 45 \cdot (0.8)^2 \cdot (0.2)^8 \\
&= 45 \cdot 0.64 \cdot (0.00000256) \\
NP_{(8)} &= 0.000737 \\
NP_{(9)} &= 10 C_9 \cdot \left(\frac{4}{5}\right)^{10-9} \cdot \left(\frac{1}{5}\right)^9 \\
&= 10 \cdot (0.8) \cdot (0.2)^9 \\
&= 10 \cdot (0.8) \cdot (0.000000512) \\
NP_{(9)} &= 0.000004096 \\
NP_{(10)} &= 10 C_{10} \cdot \left(\frac{4}{5}\right)^{10-10} \cdot \left(\frac{1}{5}\right)^{10} \\
&= 1.1 \cdot 0.0000001024 \\
NP_{(10)} &= 0.0000001024 \\
NP_{(7)} + NP_{(8)} + NP_{(9)} + NP_{(10)} &= 0.000786 + 0.0000737 + 0.000004096 + 0.0000001024 \\
&= 0.000864
\end{aligned}$$

Probability of at least 7 workers catching hold.

II Poisson Distribution

French Mathematician Simeon.D. Denis Poisson by 1837.

Uses of Poisson distribution

1. This is used in quality control to count the no. of defect of items.
2. It is used in Biology lab to count the no. of Bacteria.
3. It is used in physic lab to count no. of particles emitted from a radio active substance.
4. It is used in insurance problem to count casualties.
5. It is used in waiting time problem to the no. of incoming the customer (or) incoming telephone calls.
6. It is used in no. of type graphical error per page in typed material.
7. It is used to determine the no. of deaths occurring in α district during a given period by a rare disease.
8. It is used to count year no. of traffic and arrival such as airplanes at airport, trucks at terminals.

Poisson distribution

$$\bar{X} = m; \sigma = \sqrt{m}, p(r) = \frac{e^{-m} \cdot m^r}{r!}$$

$$m = np \quad NP_{(r)} = \frac{N \cdot e^{-m} \cdot m^r}{r!}$$

$$e = 2.7183, \log e = 0.4343$$

$$e^{-m} = \frac{1}{e^m}$$

$$P(0) = \frac{e^{-m} \cdot m^1}{1!}; \quad P(0) = e^{-m}$$

$$P(1) = \frac{e^{-m} \cdot m^1}{1!}; \quad P(1) = P(0) \cdot m$$

$$P(2) = P(1) \cdot \frac{m}{2}$$

$$P(3) = P(2) \cdot \frac{m}{3}$$

$$P(4) = P(3) \cdot \frac{m}{4}$$

$$P(5) = P(4) \cdot \frac{m}{5}$$

1. The mean of a Poisson distribution is 4. Find standard deviation.

i) β_1

ii) β_2

$$\beta_1 = \frac{1}{m} \quad \beta_2 = 3 + \frac{1}{m}$$

$$\bar{X} = m = 4$$

$$\beta_1 = \frac{1}{m} = \frac{1}{4} = 0.25$$

$$\beta_2 = 3 + \frac{1}{4} = \frac{13}{4} = 3.25$$

2. Suppose an average one house in 1000. In a certain district has a faire during the year. It there are 2000 houses in the district, what is the probability those exactly 5 houses will have a fair during the year.

$$m = np \quad p = \frac{1}{1000}; n = 2000 \quad r = 5$$

$$m = 2$$

$$P(r) = \frac{e^{-m} \cdot m^r}{r!}$$

$$P(5) = \frac{e^{-2} \cdot m^5}{5!}$$

$$= \frac{(2.7183)^{-2} \cdot 3^2}{120}$$

$$= \frac{(0.1353) \cdot 3^2}{120}$$

$$= \frac{4.3296}{120} = 0.0361$$

$$e^{-2} = (2.7183)^{-2}$$

$$= \frac{1}{(2.7183)^2}$$

$$= \frac{1}{7.3892} = 0.1353$$

3. As a tailor prepared shirts for a no. of customers after preliminary trails and retrials he had alter the shirt as follows.

No. of trails after the preliminary trails	0	1	2	3
Frequency	200	75	20	5

By applying Poisson distribution calculate the theoretical frequencies.

<i>x</i>	<i>f</i>	<i>fx</i>
0	200	0
1	75	75

2	20	40
3	5	15
N =	300	130

$$N = \Sigma F$$

$$\bar{X} = m = \frac{\Sigma fx}{N}$$

$$= \frac{130}{300}$$

$$m = \bar{X} = 0.43$$

$$NP_{(r)} = N \cdot e^{-m}$$

$$NP_{(0)} = N \times e^{-m}$$

$$NP_{(1)} = NP_{(0)} \cdot \frac{m}{1}$$

$$NP_{(2)} = NP_{(1)} \cdot \frac{m}{2}$$

$$NP_{(0)} = N \cdot e^{-m}$$

$$= 300 \cdot (2.7183)^{-0.43}$$

$$= 300 \cdot \frac{1}{(2.7183)^{0.43}}$$

$$= \text{Antilog} [\log 1 - .43 \log 2.7183]$$

$$= \text{Antilog} [0 - (0.43) \cdot (0.4343)]$$

$$= \text{Antilog} [0 - .1867]$$

$$+1 - 0.1867 - 1$$

$$0.8133 - 1$$

$$= \text{Anti log} [\bar{1}.8133]$$

$$= .6506$$

$$NP_{(0)} = N \times e^{-m} = 300 \times .6506 = 195.18$$

$$NP_{(1)} = NP_{(0)} \cdot \frac{m}{1} = 195.18 \cdot \left(\frac{.43}{1}\right) = 83.9274$$

$$NP_{(2)} = NP_{(1)} \cdot \frac{m}{2} = 83.9274 \times \left(\frac{.43}{2}\right) =$$

18.044391

$$NP_{(3)} = NP_{(2)} \cdot \frac{m}{3} = 18.044 \times \left(\frac{.43}{3}\right) = 2.586$$

4. A typist in a company commits the following no. of mistake per page in typing 432 pages.

Mistake per page (x)	0	1	2	3	4
No. of pages (F)	223	142	48	15	4

x	f	fx
0	223	0
1	142	142
2	48	96
3	15	45
4	4	16
	432	299

$$N = \Sigma F$$

$$\bar{X} = m = \frac{\Sigma fx}{N}$$

$$= \frac{299}{432}$$

$$m = 0.692$$

$$NP_{(r)} = N \cdot e^{-m}$$

$$NP_{(0)} = 432 \cdot (2.7183)^{-0.692}$$

$$= 432 \cdot \frac{1}{(2.7183)^{0.692}}$$

$$= \text{Anti log} [\log 1 - 0.692 \log 2.7183]$$

$$= \text{Anti log} [0 - (0.692) \cdot (0.4343)]$$

$$= \text{Anti log} [0 - 0.3005]$$

$$= \text{Anti log} [\bar{1}. (0.6995)]$$

$$= 0.5006$$

$$NP_{(0)} = N \cdot e^{-m} = 432 \times (0.5006) = 216.2592$$

$$NP_{(1)} = NP_{(0)} \cdot \frac{m}{1} = 216.259 \times \frac{0.692}{1} = 149.651$$

$$NP_{(2)} = NP_{(1)} \cdot \frac{m}{2} = 149.651 \times \frac{0.692}{2} = 51.779$$

$$NP_{(3)} = NP_{(2)} \cdot \frac{m}{3} = 51.779 \times \frac{0.692}{3} = 11.944$$

$$NP_{(4)} = NP_{(3)} \cdot \frac{m}{4} = 11.944 \times \frac{0.692}{4} = 2.066$$

$$NP_{(0)} + NP_{(1)} + NP_{(2)} + NP_{(3)} + NP_{(4)} = 431.6992$$

1. In a manufacturing organization the distribution of wages was normal and the no. of workers employed in the organization was 5000, the mean wages of workers were calculate at Rs. 800 (p.m) and

σ worked out to Rs. 200 Rs. on the basis of information estimate. i) The no. of workers getting salary between Rs.700 and Rs.900. ii) The percentage of worker getting salary above Rs.1000. iii) the percentage of worker getting salary below Rs.600.

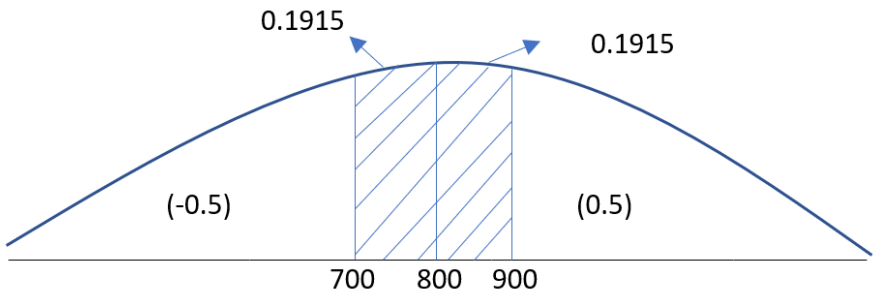
$$i) Z = \frac{x - \bar{x}}{\sigma} \qquad \bar{X} = 800, \sigma = 200 \quad x = 700 \text{ and } 900$$

$$= \frac{700 - 800}{200}$$

$$Z = -0.5$$

$$= \frac{900 - 800}{200}$$

$$Z = 0.5$$



The no. of workers getting salary between Rs. 700 and Rs.900 is $= 0.1915 + 0.1915 = \sqrt{0.383}$.

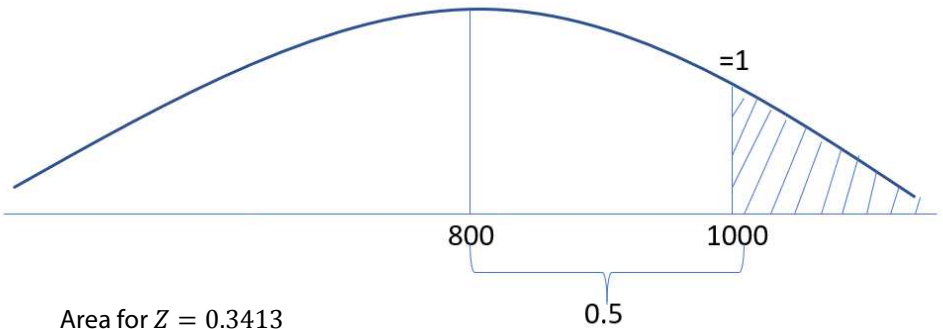
Area for $Z = 0.5$ and the table value is 0.1915 another area for $Z = -0.5$ and the table value is 0.1915.

Required area is $0.1915 + 0.1915 = 0.3830$.

No. of workers required area $= 5000 \text{ workers} \times 0.3830 = 1915$.

$$ii) Z = \frac{1000 - 800}{200}$$

$$Z = 1$$

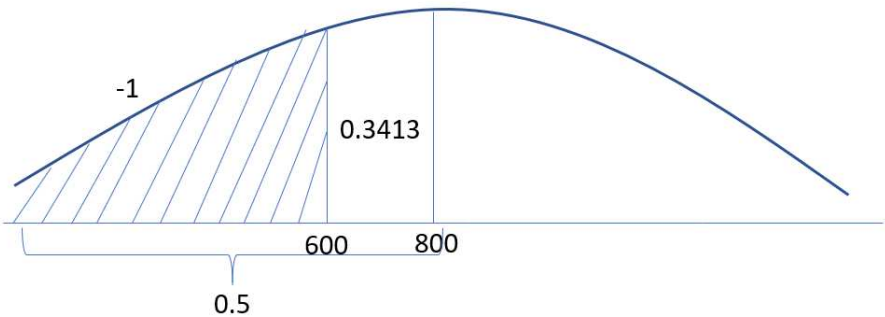


Area for $Z = 0.3413$

Required area is $=0.5-0.3143=0.1587$

Percentage of workers $= 0.1587 \times 100 = 15.87\%$

iii)
$$Z = \frac{600-800}{200} = \frac{-200}{200} = -1$$



Area for $(+0.3413)$

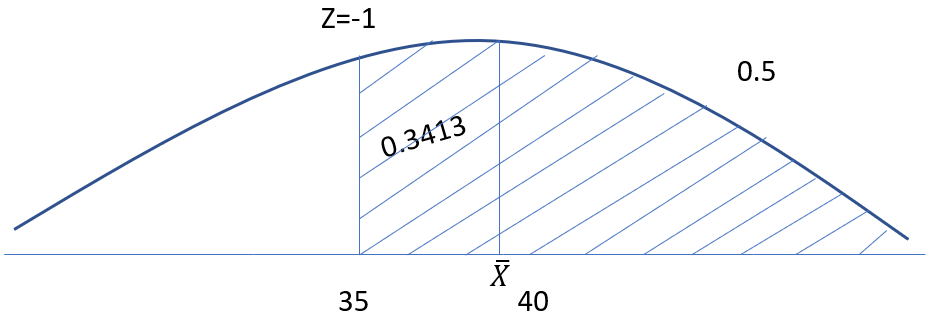
Required area is $= -0.5 - 0.3413 = 0.1587$

Percentage of workers $= 0.1587 \times 100 = 15.87\%$

2. Suppose, that the life of the gas cylinder is normally distributed the mean of 40 days, σ of 5 days, if at a time 10000 cylinder are issued to customers how many will need after 35 days.

$x = 35 \text{ days}, \bar{X} = 40 \text{ days}, \sigma = 5 \text{ days}$

$$Z = \frac{x-\bar{x}}{\sigma} = \frac{35-40}{5} = \frac{-5}{5} = -1$$



Area for $=0.3413$

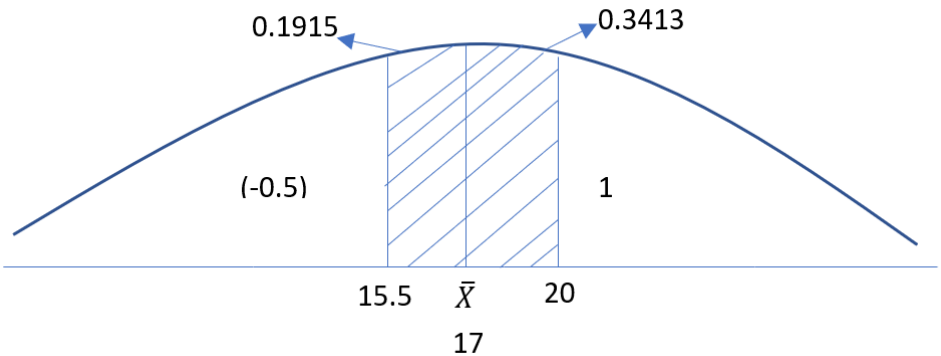
Required area is $=0.5+0.3413 = 0.8413$

No. of customers requires repayment after 35 days $= 0.8413 \times 10000 = 8413$.

3. The mean yield per plot of a cloth in 17kg and σ is 3 kg the distribution of yield per plot is normal. Find there percentage of plot living yield.

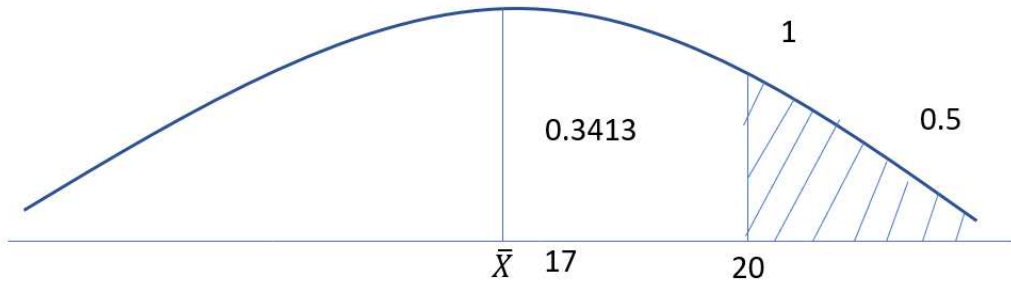
i) Between 15.5kg and 20kg ii) more than 20kg.

$$\begin{aligned}
 \text{i) } Z &= \frac{x-\bar{X}}{\sigma} & \bar{X} &= 17kg, \sigma = 3kg \\
 &= \frac{15.5-17}{3} \\
 &= -0.5 \\
 Z &= \frac{20-17}{3} = 1.
 \end{aligned}$$



i) Between 15.5 kg and 20kg $= 0.1915+0.3413=0.5328$.

ii) $Z = \frac{20-17}{3} = 1.$



More than 20kg is $= 0.5 - 0.3413 = 0.1587$

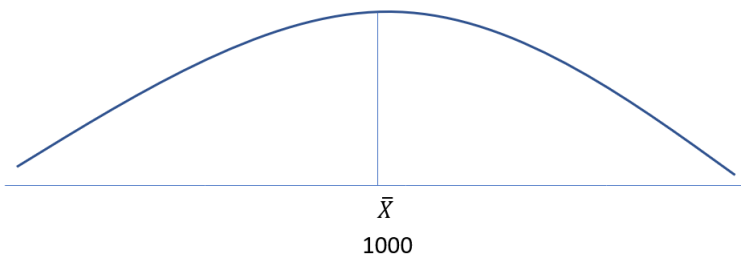
5. A Machine fills cotton such that a mean of the distribution of fills is 1 ltr is the σ of 10 milk. If the distribution is normal

- i) What percentage of cotton will contains less than $\frac{1}{2}$ lt.
- ii) What percentage of cotton will contains less than 975 m. ltr.
- iii) More than 1010 m. ltr.

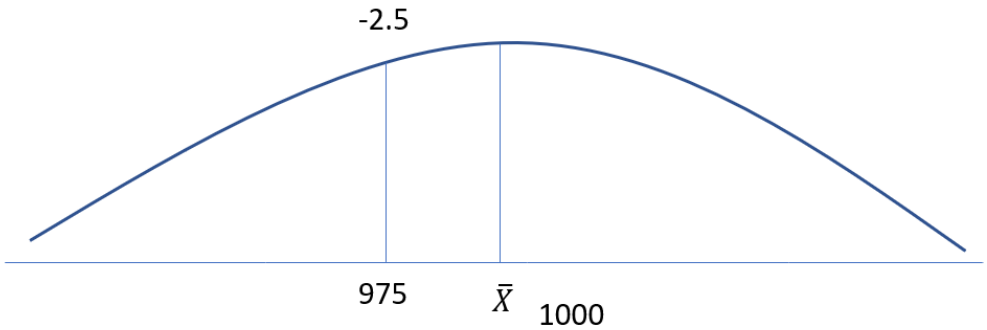
1 Liter = 1000 m.ltr.

$$\bar{x} = 1000, \sigma = 10$$

i. $Z = \frac{x-\bar{x}}{\sigma} = \frac{1000-1000}{10} = 0$

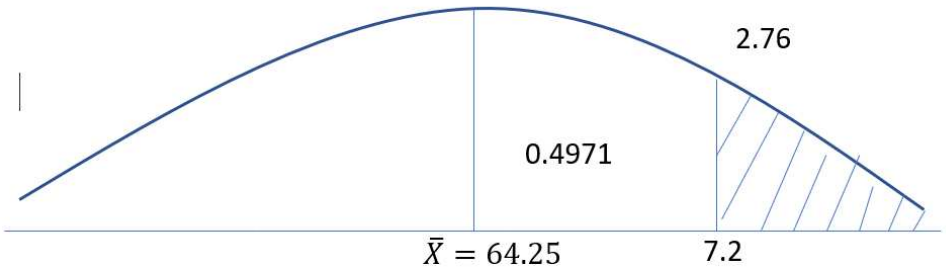


$$\text{ii. } Z = \frac{x - \bar{x}}{\sigma} = \frac{975 - 1000}{10} = -2.5$$



5. Assume that the mean height of the group of men is 64.25 inches to the σ 2.81 inches. How many in the group of 400 men would you expect to be over 6 feet of? 1 feet = 12 inches, (6 x 12) feet = 72 inches

$$Z = \frac{x - \bar{x}}{\sigma} = \frac{72 - 64.25}{2.81} = 2.76$$



$$0.5 - 0.4971 = 0.0029$$

$$\text{No. of feed} = 400 \text{ man} \times 0.0029 = 1.16$$

$$\begin{array}{r} 280 = 0.4974 \\ (-) 275 = 0.4970 \\ \hline 0.25 = 0.0004 \end{array}$$

$$\frac{0.05}{5} = 0.01$$

$$\frac{0.0004}{0.05} \times 0.01 = 0.00008$$

	Z		Table
	2.75	=	0.4970
(+)	0.01	=	0.00008
	2.76	=	0.49708

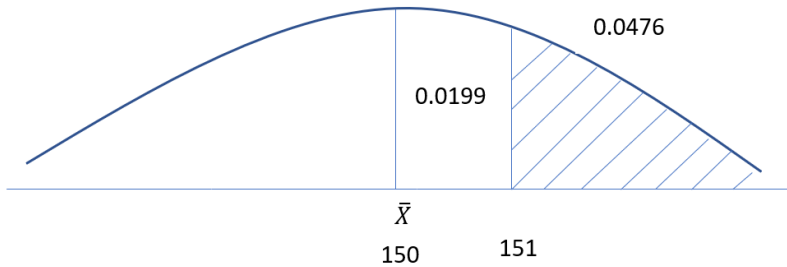
$$Z=0.4971$$

6. The large population at a mean height of 150cm and σ 21 cm. A random sample of size is 100 taken this population.

- i) Find the probability that the sample mean will exceed 151 cm.
- ii) lice between 148 cm and 155 cm.

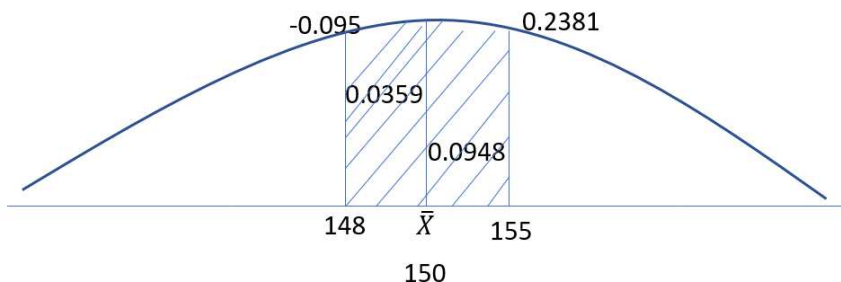
Assume that the population distribution of height is normal and the sampling with replacement.

$$i) \quad Z = \frac{x - \bar{x}}{\sigma} = \frac{151 - 150}{21} = 0.0476$$



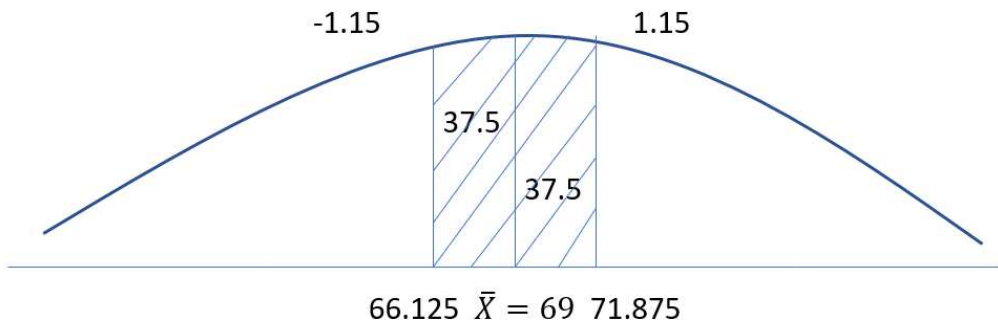
$$0.5 - 0.0199 = 0.4801$$

$$ii) \quad Z = \frac{x - \bar{x}}{\sigma} = \frac{148 - 150}{21} = -0.095 \quad \quad Z = \frac{x - \bar{x}}{\sigma} = \frac{155 - 150}{21} = 0.2381$$



$$\begin{aligned} \text{Between 148 cm and 155 cm} &= 0.0359 + 0.0948 \\ &= 0.1307 \end{aligned}$$

7. The height of 12000 college student closely follows a normal distribution with mean 69 inch and σ 2.5 inches. What range of height would you expect to include the middle 75% of the students in this group?



$$\text{Area of table value} = \frac{37.5}{100} = 0.375$$

$$\therefore Z = 1.15$$

$$Z = \frac{X - \bar{X}}{\sigma}$$

$$1.15 = \frac{X - 69}{2.5}$$

$$1.15 \times 2.5 = X - 69$$

$$2.875 = X - 69$$

$$69 + 2.875 = X$$

$$71.875 = X$$

$$Z = \frac{X - \bar{X}}{\sigma}$$

$$-1.15 = \frac{X - 69}{2.5}$$

$$-1.15 \times 2.5 = X - 69$$

$$-2.875 = X - 69$$

$$69 - 2.875 = X$$

$$66.125 = X$$

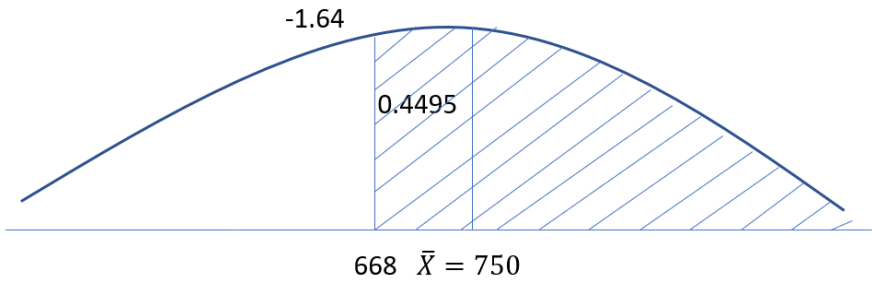
8. The income of groups of 10000 person was count to be normally distributed with mean 750 per month and σ 50.

Find i) the percentage of the person having income exceeding Rs.668.

ii) The No. of person having income exceeds Rs. 832.

iii) What was the lowest income among the richest Rs.100?

$$Z = \frac{x - \bar{x}}{\sigma} = \frac{668 - 750}{50} = -1.64$$



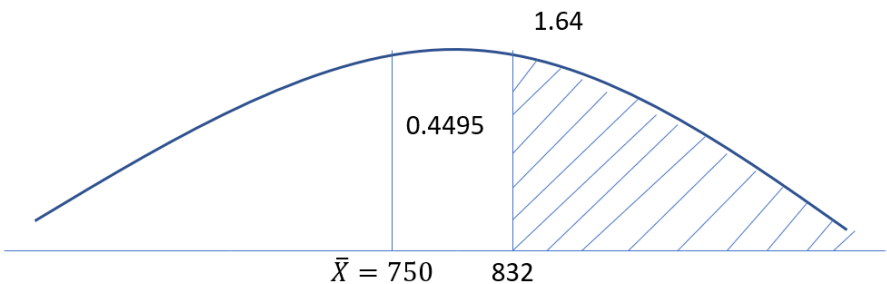
Total area required

$$= 0.5 + 0.4495 = 0.9495 \times 100$$

$$= 94.95\%$$

% of the person having income exceeding Rs.668 = 94.95%

iii) $Z = \frac{x - \bar{x}}{\sigma} = \frac{832 - 750}{50} = 1.64$



Total area required

$$= 0.5 - 0.4495 = 0.0505$$

No. of person having income exceeds Rs.832 = $0.0505 \times 10000 = 505$

$$\text{iii) } \frac{100}{1000} = 0.01$$

$$0.5 - 0.01 = 0.49$$

Table value = 0.4901

$$\therefore Z = 2.33$$

$$Z = \frac{x - \bar{x}}{\sigma}$$

$$2.33 = \frac{x - 750}{50}$$

$$2.33 \times 50 = x - 750$$

$$116.5 = x - 750$$

$$116.5 + 750 = x$$

$$866.5 = x$$

a) Fit a normal curve with a following data by the method of area.

Variable	Frequency
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

Solution:

Variable	Mind value	Frequency	d' $= \frac{m - A}{C}$	Fd'
59.5-62.5	61	5	-2	-10
62.5-65.5	64	18	-1	-18
65.5-68.5	67	42	0	0
68.5-71.5	70	27	1	27
71.5-74.5	73	8	2	16
		N=100		15

$$\bar{X} = A + \frac{\sum d^1}{N} \times C$$

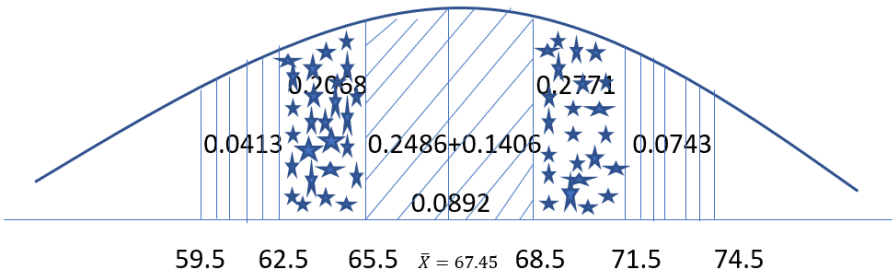
$$\begin{aligned}
 &= 67 + \frac{15}{100} \times 3 \\
 &= 67 + 0.15 \times 3 \\
 &= 67 + 0.45 \\
 &= 67.45
 \end{aligned}$$

$$\begin{aligned}
 \sigma &= \sqrt{\left(\frac{\sum Fd^2}{N}\right) - \left(\frac{\sum Fd}{N}\right)^2} \times C \\
 &= \sqrt{\frac{97}{100} - \left(\frac{15}{100}\right)^2} \times 3 \\
 &= \sqrt{0.97 - (0.15)^2} \times 3 \\
 &= \sqrt{0.97 - 0.0225} \times 3 \\
 &= \sqrt{0.9475} \times 3 \\
 &= 0.9734 \times 3 \\
 \sigma &= 2.9202
 \end{aligned}$$

Mid value	F	$d' = \frac{m - A}{C}$	d^2	Fd	Fd^2
61	5	-2	4	-10	20
64	18	-1	1	-18	18
67	42	0	0	0	0
70	27	1	1	27	27
73	8	2	4	16	32
	N=100			15	97

Lower List	$Z = \frac{x - \bar{x}}{\sigma}$	Area	Class interval	Area required	Theoretical Frequency
59.5	-2.722	0.4967	59.5-62.5	0.0413	4.13
62.5	-1.70	0.4554	62.5-65.5	0.2068	20.68
65.5	-0.6678	0.2486	65.5-68.5	0.3892	38.92
68.5	0.3596	0.1406	68.5-71.5	0.2771	27.71

71.5	1.3869	0.4177	71.5-74.5	0.0743	7.43
74.5	2.4142	0.4920	-	-	100



2.72 - Table value

	Z	=	Table
	2.75	=	0.4970
(-)	2.70	=	0.4965
	<u>0.05</u>	=	<u>0.0005</u>
	<u>5</u>		<u>5</u>
	=0.01		=0.0001

$$\frac{0.0001}{0.01} \times 0.02 = 0.0002$$

	Z	=	Table
	2.70	=	0.4965
(+)	0.02	=	0.0002
	<u>2.72</u>	=	<u>0.4967</u>

1.70 - Table value

0.67 - Table value

	Z	=	Table
	0.70	=	0.2580
(-)	0.65	=	0.2422
	<u>0.05</u>	=	<u>0.0158</u>
	<u>5</u>		<u>5</u>

=0.01 =0.00316

$$\frac{0.00316}{0.01} \times 0.02 = 0.00632$$

	Z	=	Table
	0.65	=	0.2422
(+)	0.02	=	0.00632
	<u>0.67</u>	=	<u>0.24852</u>

0.36 - Table value

	Z	=	Table
	0.40	=	0.1554
(-)	0.35	=	0.1368
	<u>0.05</u>	=	<u>0.0186</u>
	<u>5</u>		<u>5</u>

=0.01 =0.00372

	Z	=	Table
	0.35	=	0.1368
(+)	0.01	=	0.00372
	<u>0.36</u>	=	<u>0.14052</u>

1.39 - Table value

	Z	=	Table
	40	=	0.4192
(-)	1.35	=	0.4115
	<u>0.05</u>	=	<u>0.0077</u>
	<u>5</u>		<u>5</u>

=0.01 =0.00154

$\frac{0.00154}{0.01} \times 0.04 = 0.00616$

	Z	=	Table
	1.35	=	0.4115
(+)	0.04	=	0.00616
	<u>1.39</u>	=	<u>0.41766</u>

2.41 - Table value

	Z	=	Table
	2.45	=	0.4929
(-)	2.40	=	0.4918
	<u>0.05</u>	=	<u>0.0011</u>

=0.01 =0.00022

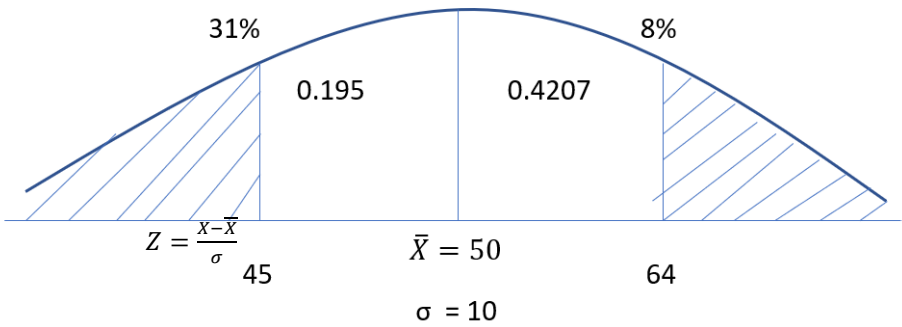
$\frac{0.00154}{0.01} \times 0.04 = 0.00616$

	Z	=	Table
	2.40	=	0.4918
(+)	0.01	=	0.00022
	<u>2.41</u>	=	<u>0.49202</u>

10. Find the mean and σ of a normal distribution when 8% of the item are over 64 and 31% (or) under 45 of a large group of men 33% of under 165 cm in high and 63% or between 165 × 185 cm. Find out the arithmetic mean and σ of the distribution of height. Assume is to be normal.

i) $X = 64 ; X = 45$

8%	31%
$= \frac{8}{100} = 0.08$	$= \frac{31}{100} = 0.31$
Area = 0.4207	Area = 0.1915
$Z = 1.41$	$Z = 0.50$



$$-0.50 = \frac{45 - \bar{X}}{\sigma}$$

$$-0.50\sigma = 45 - \bar{X} \quad \text{---}$$

(1)

$$Z = \frac{x - \bar{X}}{\sigma}$$

$$-1.41 = \frac{64 - \bar{X}}{\sigma}$$

$$1.41\sigma = 64 - \bar{X} \quad \text{---}$$

(2)

$$-0.50\sigma = 45 \quad \text{--- (1)}$$

$$+1.41\sigma = 64 \quad \text{--- (2)}$$

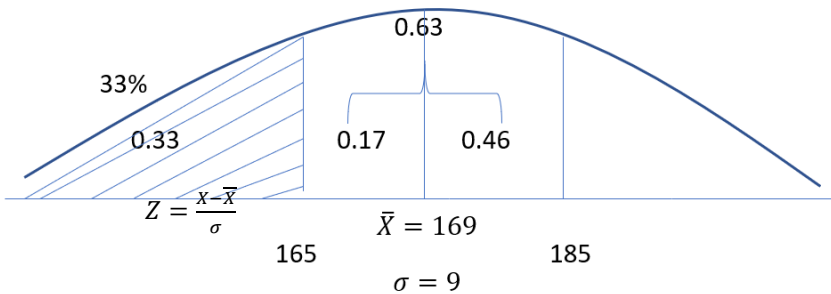
$$-1.91\sigma = -19$$

$$\sigma = \frac{-19}{-1.91}$$

$$\begin{aligned} \sigma &= 9.948 \\ \sigma &= 10 \\ 1.41 \sigma &= 64 - \bar{X} \\ 1.41 (9.948) &= 64 - \bar{X} \\ 14.03 &= 64 - \bar{X} \\ -64 + 14.03 &= -\bar{X} \\ 49.97 &= \bar{X} \\ \bar{X} &= 49.97 = 50 \end{aligned}$$

ii) $X = 165$; $X = 165$ & 185

$$\begin{array}{llll} 33\% & = & 0.33 & \quad \quad \quad 63\% & = & 0.63 \\ (0.5-0.33) & & & & (0.63-0.17) & & \\ \text{Area} = 0.1700 & & & & \text{Area} = 0.4599 & & \\ \mathbf{Z=0.44} & & & & \mathbf{Z=1.75} & & \end{array}$$



$$\begin{aligned} -0.44 &= \frac{165 - \bar{X}}{\sigma} \\ -0.44\sigma &= 165 - \bar{X} \quad \quad \quad \dots \quad (1) \end{aligned}$$

$$\begin{aligned} Z &= \frac{x - \bar{X}}{\sigma} \\ 1.75 &= \frac{185 - \bar{X}}{\sigma} \\ 1.75\sigma &= 185 - \bar{X} \quad \quad \quad \dots \\ (2) & \end{aligned}$$

$$-0.44 \sigma = 165 - \bar{X} \quad \text{--- (1)}$$

$$+1.75 \sigma = 185 - \bar{X} \quad \text{--- (2)}$$

$$-2.19 \sigma = -20$$

$$\sigma = \frac{-20}{-2.19}$$

$$\sigma = 9.13$$

$$1.75\sigma = 185 - \bar{X} \quad \text{---}$$

(2)

$$1.75(9.13) = 185 - \bar{X}$$

$$15.978 = 185 - \bar{X}$$

$$\bar{X} = 185 - 15.978$$

$$\bar{X} = 169.02$$

Chapter – II

ANOVA Table

AN = Analysis

O = Of

VA = Variance

1. One Way Classification
2. Two Way Classification

One Way Classification:-

Source of Variance	Sum of Squares	Degree of Freedom	Mean Square	Variance
Between Samples	SSC	U=c-1	MSC = $\frac{SSC}{c-1}$	$F = \frac{MSE}{MSE}$
Within Samples	SSE	U=N-C	MSE = $\frac{SSE}{N-C}$	
Total	SST	N-1	-	

1) Correction Total = $\frac{T^2}{N}$
 $T = \sum x_1 + \sum x_2 + \sum x_3 + \sum x_4$

2) Sum of Squares Total (SST)
 $SST = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2 - \frac{T^2}{N}$ (C.T)

3) Sum of Squares between Column or Samples
 $SSC = \frac{(\sum x_1)^2}{N} + \frac{(\sum x_2)^2}{N} + \frac{(\sum x_3)^2}{N} + \frac{(\sum x_4)^2}{N} - \frac{T^2}{N}$ (C.T)

SST = Sum of Squares Total

SSC = Sum of Squares Between Columns or Samples

SSE = Sum of Squares within samples

CT = Correction Total (Correction factor)

MSC = Mean Square between Column

MSE = Mean Square within Samples

Meaning – ANOVA

The analysis of variance is a statistical technique specially designed to test whether a means of square more than 2 quantitative population are equal.

The ANOVA originated in agrarian research and its language is thus loaded with such agricultural terms block (referring to learn and treatment referring to population) which are different terms of varieties of seeds, fertilizer or cultivation method.

Assumption of ANOVA:

1. Normality
2. Homogeneity
3. Independence of Error

1) To assess the significance of possible variance in the performance in a certain test was given to a No.of students taken at a random from the senior fifth class of each. The four school concern.

The results are given below make an analysis of variance of data

A	B	C	D
x1	x2	x3	x4
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15
45	50	60	65

2) 3 samples below have been obtain from a normal population with equal variance. Test the hypothesis that the sample mean was equal.

A	B	C
8	7	12
10	5	09
7	10	13
14	9	12
11	9	14

3) The following table gives monthly sales in Rs.1,000 if a certain firm in 3 states by 4 salesman.

Status	Salesman				Total
	I	II	III	IV	
A	6	5	3	8	22
B	8	9	6	5	28
C	10	7	8	7	32
Total	24	21	17	20	82

Set up the analysis of variance table and list whether there is any significance difference among 4 salesman within the 3 states.

1st Problem Conclusion

Table value of variance $F_{0.05} \sqrt{1} = 3$ and $\sqrt{12} = 16$ table value 3.2 ($1.28 < 3.24$). Since the call value F is less than the table value f 5%. 5% level of significance. Our hypothesis is accepted. Hence we conclude that there is no significant difference among the means of different 4 school.

ANOVA (One Way Classification)

1)

A		B		C		D	
8	64	12	144	18	324	13	169
10	100	11	121	12	144	9	81
12	144	9	81	16	256	12	144
8	64	14	196	6	36	16	256
7	49	4	16	8	64	15	225
45	421	50	558	60	824	64	875

$$SST = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2 - \frac{T^2}{N}$$

$$\frac{T^2}{N} \rightarrow T = \sum x_1 + \sum x_2 + \sum x_3 + \sum x_4$$

$$T = 45 + 50 + 60 + 65$$

$$\begin{aligned} \frac{T^2}{N} &= \frac{220^2}{20} = \frac{48400}{20} \\ &= 2420 \end{aligned}$$

$$SST = 421 + 558 + 824 + 875 - 2420$$

$$SST = 258$$

$$\begin{aligned}
 SSC &= \frac{(\sum x_1)^2}{N} + \frac{(\sum x_2)^2}{N} + \frac{(\sum x_3)^2}{N} + \frac{(\sum x_4)^2}{N} - \frac{T^2}{N} \\
 &= \frac{45^2}{5} + \frac{50^2}{5} + \frac{60^2}{5} + \frac{65^2}{5} - 2420 \\
 &= 2470 - 2420
 \end{aligned}$$

$$SSC = 50$$

Hypothesis

Let us take the hypothesis that there is no significant difference among the means of different 4 school.

ANOVA Table

Sources of Variance	Sum of Square	Degree of Freedom	Mean Square	Variance
Between Sample	SSC 50	= C - 1 4-1 = 3	MSC = $\frac{SSC}{C-1}$ = 50/3 = 16.67	F = $\frac{MSC}{MSE}$
Within Sample	SSE 208	= N - C 20 - 4 = 16	MSE = $\frac{SSE}{N-C}$ = 208/16 = 13	= 16.67/13
Total	SST 258	N - 1 20 - 1 = 19	-	= 1.28

2)

A		B		C	
8	64	7	49	12	144
10	100	5	25	9	81
7	49	10	100	13	169
14	196	9	81	12	144
11	121	9	81	14	196
50	530	40	336	60	734

$$\begin{aligned}
 SST &= \sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2 - \frac{T^2}{N} \\
 \frac{T^2}{N} &\Rightarrow \sum x_1 + \sum x_2 + \sum x_3 + \sum x_4 \\
 &= 50+40+60 \\
 &= 150/15 = 22500/15
 \end{aligned}$$

$$T^2/N = 1500$$

$$\text{SST} = 530 + 336 + 734 - 1500$$

$$= 1600 - 1500$$

$$\text{SST} = 100$$

$$\text{SSC} = \frac{(\sum x_1)^2}{N} + \frac{(\sum x_2)^2}{N} + \frac{(\sum x_3)^2}{N} + \frac{(\sum x_4)^2}{N} - T^2/N$$

$$= \frac{50^2}{5} + \frac{40^2}{5} + \frac{60^2}{5} - 1500$$

$$= \frac{2500}{5} + \frac{1600}{5} + \frac{3600}{5} - 1500$$

$$= 500 + 320 + 720 - 1500$$

$$\text{SSC} = 40$$

Source of Variance	Sum of Square	Degree of freedom	Mean Square	Variance
Between Sample	SSC = 40	= C - 1 = 3 - 1 = 2	$40/2 = 20$	
Within Sample	SSE = 60	= N - C = 15 - 3 = 12	$60/12 = 5$	
Total	SST = 100	= N - 1 = 15 - 1 = 14	-	$F = 20/5 = 4$

Two Way Classification

Degree of Variance	Sum of Squares	Degree of freedom	Mean Square	Variance
Between Samples (Columns)	SSC	C-1	$MSC = \frac{SSC}{C-1}$	$F_1 = \frac{MSC}{MSE}$
Between States (Rows)	SSR	r-1	$MSR = \frac{SSR}{r-1}$	$F_2 = \frac{MSR}{MSE}$
Residual Error	SSE	(c-1)(r-1)	$\frac{MSE = SSE}{(c-1)(r-1)}$	
Total	SST	N-1	-	

$$SSR = \frac{(\sum y_1)^2}{N} + \frac{(\sum y_2)^2}{N} + \frac{(\sum y_3)^2}{N} - T^2/N$$

2) Hypothesis

Let us take the hypothesis that the sample means are equal

Conclusion

Since the calculated value F is greater than 5% significant level.

Table value (4 > 3.88). Hence hypothesis is rejected. The sample means are not equal.

Two Way Classification

Hypothesis

Let us take the hypothesis that there is no significant difference among the salesman and there is no significant different among the states in 3 states.

States	I		II		III		IV		Total
A	6	36	5	25	3	9	8	64	22
B	8	64	9	81	6	36	5	25	28
C	10	100	7	49	8	64	7	49	32
Total	24	200	21	155	17	109	20	138	82

$$SST = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2 - T^2/N$$

$$T^2/N \quad \begin{array}{l} T = \sum x_1 + \sum x_2 + \sum x_3 + \sum x_4 \\ \rightarrow = 24 + 21 + 17 + 20 \\ = 150/15 = 560.33 \end{array}$$

$$SST = 200 + 155 + 109 + 138 - 560.33 = 602 - 560.33$$

$$SST = 41.67$$

$$SSC = \frac{(\sum x_1)^2}{N} + \frac{(\sum x_2)^2}{N} + \frac{(\sum x_3)^2}{N} + \frac{(\sum x_4)^2}{N} - T^2/N$$

$$= \frac{24^2}{3} + \frac{21^2}{3} + \frac{17^2}{3} + \frac{20^2}{3} - 560.33$$

$$= 576/3 + 441/3 + 289/3 + 400/3 - 560.33$$

$$= 192 + 147 + 96.33 + 133.33 - 560.33$$

$$SSC = 8.33$$

$$SSR = \frac{(\sum y_1)^2}{N} + \frac{(\sum y_2)^2}{N} + \frac{(\sum y_3)^2}{N} - T^2/N$$

$$\begin{aligned}
&= \frac{22^2}{3} + \frac{28^2}{3} + \frac{32^2}{3} - 560.33 \\
&= 121 + 196 + 256 - 560.33 \\
&= 573 - 560.33 = 12.67
\end{aligned}$$

Two Way Classification

Degree of Variance	Sum of Squares	Degree of freedom	Mean Square	Variance
Between Samples (Columns)	SSC = 8.33	C-1 = 4-1 = 3	$MSC = \frac{SSC}{C-1}$ $= \frac{8.33}{3} = 2.77$	$F_1 = \frac{MSC}{MSE}$ $= \frac{2.77}{3.445} = 0.804$
Between States (Rows)	SSR = 12.67	r-1 = 3-1 = 2	$MSR = \frac{SSR}{r-1}$ $= \frac{12.67}{2} = 6.335$	$F_2 = \frac{MSR}{MSE} = \frac{6.335}{3.445} = 1.839$
Residual Error	SSE = 20.67	(c-1)(r-1) = 3 x 2 = 6	$MSE = \frac{SSE}{(c-1)(r-1)}$ $= \frac{20.67}{6} = 3.445$	
Total	SST = 41.67	N-1 = 12-1 = 11	-	

Conclusion:

Table value of Variance (f)

D.F (3,6)

F 0.05 – 4.76 Table Value of F 0.06

2,6 = 5.14

Since the calculated value of F1 is less than the table value of 5% level of significance. (0.807 is < 4.76). The hypothesis is accepted.

Hence there is no significance differences among the sales by the firm of 4 salesman.

2) F2 calculated value is less than the table value at 5% level of significance. (1.839 < 5.14).

This hypothesis is accepted. There is no significance difference among the sales in 3 states.

Coding Table value of Variance (f)

D.F (3,6)

F 0.05 – 4.76 Table Value of F 0.06

$$2,6 = 5.14$$

Since the calculated value of F1 is less than the table value of 5% level of significance. (0.807 is < 4.76). The hypothesis is accepted.

Hence there is no significance differences among the sales by the firm of 4 salesman.

2) F2 calculated value is less than the table value at 5% level of significance. (1.839 < 5.14).

This hypothesis is accepted. There is no significance difference among the sales in 3 states.

Coding

- 1) A manufacturing company has purchase 3 new machines of different production and wishes to determine whether one of them is faster than the others. In producing a certain output. 5 hourly production figures are observed at random from each machine and the results are given below.

A ₁	A ₂	A ₃
25	31	24
30	39	30
36	38	28
38	42	25
31	35	28

Use analysis of variance and determine whether the machines significant difference in speen. Table F 0.05 – D.F (2,12) = 3.88

A ₁	A ₂	A ₃
\bar{x}_1	\bar{x}_2	\bar{x}_3
-5	1	-6
0	9	0
6	8	-2
8	12	-5
1	5	-2

Hypothesis

Let us take the hypothesis that there is no significant difference among the 3 machine speed.

$$\begin{aligned} T &= \bar{x}_1 + \bar{x}_2 + \bar{x}_3 \\ &= 10 + 35 + (-15) \\ &= 30 \end{aligned}$$

$$\frac{\bar{x}^2}{n} = \frac{30^2}{15} = \frac{900}{15} = 60$$

$$\begin{aligned} SST &= \sum \bar{x}_1^2 + \sum \bar{x}_2^2 + \sum \bar{x}_3^2 + \sum \bar{x}_4^2 - \frac{\bar{x}^2}{n} \\ &= 126 + 315 + 69 - 60 \\ &= 450 \end{aligned}$$

$$\begin{aligned} SSC &= \frac{(\sum \bar{x}_1)^2}{n} + \frac{(\sum \bar{x}_2)^2}{n} + \frac{(\sum \bar{x}_3)^2}{n} + \frac{(\sum \bar{x}_4)^2}{n} - \frac{\bar{x}^2}{n} \\ &= \frac{10^2}{5} + \frac{35^2}{5} + \frac{(-15)^2}{5} + 60 \\ &= 20 + 245 + 45 - 60 \end{aligned}$$

$$SSC = 250$$

Sources of Variance	Sum of Squares	Degree of freedom	Mean Square	Variance
Between Samples	SSC = 250	C-1 = 3-1 = 2	MSC = $\frac{250}{2} = 125$	$F_1 = \frac{125}{16.667} = 7.499$
Within Sample	SSE = 200	= N - C = 15 - 3 = 12	$\frac{200}{12} = 16.667$	
Total	SST = 450	= N - 1 = 15 - 1 = 14	-	

Conclusion

Since the calculated value of F is greater than 5% level table value (7.5 > 3.88). Hence our hypothesis rejected. Hence the machines are significantly different in their mean speed.

2)

Sales	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	\bar{x}_1^2	\bar{x}_2^2	\bar{x}_3^2	\bar{x}_4^2
A	-20	0	-5	-30	400	0	25	900

B	-18	10	-10	-25	324	100	100	625
C	-12	0	0	-18	144	25	0	324
D	-15	5	5	-12	225	225	25	144
E	-25	-10	-20	-35	625	100	400	1225
Total	-90	20	-30	-120	1718	450	550	3218

$$T = -90 + 20 + (-30) + (-120)$$

$$T = -220$$

$$T^2/N = (-120)^2/20 = 2420$$

$$\begin{aligned} SST &= \sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2 - T^2/N \\ &= 1718 + 450 + 550 + 3218 - 2420 \end{aligned}$$

$$SST = 3516$$

2) The following figures related to the number of units sold in 5 different areas by 4 salesman.

States or Areas	1	2	3	4	Total
A	80	100	95	70	
B	82	110	90	75	
C	88	105	100	82	
D	85	115	105	88	
E	75	90	80	65	

If there is any significant difference of salesman. Also find out whether there is a significant difference among the sales made in 5 area or states.

Table value of F 0.05 5% level of significance D.F (3,12) = 3.49, F_2 D.F (4,12) = 3.26

$$\begin{aligned} SSC &= \frac{(-90)^2}{5} + \frac{20^2}{5} + \frac{(-30)^2}{5} + \frac{(-120)^2}{5} - 60 \\ &= 620 + 80 + 180 + 2880 - 2420 \end{aligned}$$

$$SSC = 2340$$

$$SSR = \frac{(\sum y_1)^2}{N} + \frac{(\sum y_2)^2}{N} + \frac{(\sum y_3)^2}{N} + \frac{(\sum y_4)^2}{N} - T^2/N$$

$$y_1 = -20 + 0 + (-5) + (-30) = -55$$

$$y_2 = -18 + 10 + (-10) + (-25) = -43$$

$$\begin{aligned}
 y_3 &= (-12) + 5 + 0 + (-18) = -25 \\
 y_4 &= (-25) + (-10) + (-20) + (-35) = -90 \\
 &= \frac{(-55)^2}{4} + \frac{(-43)^2}{4} + \frac{(-15)^2}{4} + \frac{(-7)^2}{4} + \frac{(-90)^2}{4} - 240 \\
 &= 756.25 + 462.25 + 156.25 + 12.25 + 2025 - 240 \\
 \text{SSR} &= 992
 \end{aligned}$$

Degree of Variance	Sum of Squares	Degree of freedom	Mean Square	Variance
Between Samples (Columns)	SSC = 23240	C-1 = 4-1 = 3	$MSC = \frac{SSC}{C-1}$ $= \frac{2340}{3} = 780$	$F_1 = \frac{MSC}{MSE}$ $= \frac{780}{15.33} = 50.88$
Between States (Rows)	SSR = 992	r-1 = 5-1 = 4	$MSR = \frac{SSR}{r-1}$ $= \frac{992}{4} = 248$	$F_2 = \frac{248}{15.33} = 16.177$
Residual Error	SSE = 184	(c-1)(r-1) = 3 x 4 = 12	$\frac{MSE=SSE}{(c-1)(r-1)}$ $\frac{20.67}{6} = 3.445$	
Total	SST = 3516	N-1 = 20- 1 = 19	-	

Conclusion:

Since the calculated value F_1 is $50.88 > 3.49$ (Rej) and F_2 is $16.177 > 3.26$ (Rej). Since the conclude value of $70.05 = 50.88 > 3.49$ and $70.05 = 16.177 > 3.26$ at 5% Level of Significance. Our hypothesis is rejected.

Chapter – III

CHI – SQUARE TEST

Meaning

A chi – square test (...) statistic is a test that measures how expectations compare to actual observed data (or) model results. The data used in calculating a chi – square statistic must be random, raw, mutually, exclusive, drawn from independent variables and drawn from a large enough sample.

Ex;

The results of losing a coin 100 times met these criteria. Chi square tests are often used in hypothesis testing

$$x^2 = \sum \frac{(O - E)^2}{E}$$

C = degree of freedom

D = observed value

E = expected value.

Calculation of the CHI – SQUARE test;

Step 1 : identify the characteristics and response.

Step 2 : calculate the margin summations.

Step 3 : simply sum all rows and columns and enter these sums on the 'margins'.

Complete the contingency table.

Step 4 : calculate the observed chi – square value based on contingency table.

Step 5 : use standard table to compare observed chi – square to critical value of chi – square for the specific degree of freedom and confidence level.

Uses of the chi – squared statistic;

The chi – squared statistic provides a test of the association between two or more groups, populations or criteria

The chi – square test can be used to test the strength of the association between exposure and disease in a cohort study, an unmatched case – control study or a cross sectional study.

Chi – square test can be used to test the strength of the association between exposure and disease in a matched case control study.

The χ^2 statistic is calculated under the assumption of no association.

A large value of the chi - square test statistic => small probability => support an association.

A small value of the chi – square test statistic => large probability => does not support an association between exposure and disease.

Misuse of chi – square test;

Some sources of error in the application of χ^2 test are as follows;

- Small theoretical frequencies.
- Neglect of frequencies of non – occurrences.
- Failure to equalize the sum of observed frequencies and the sum of the theoretical frequencies.
- Indeterminate theoretical frequencies.
- Incorrect or questionable categorizing,
- Use of non – frequency data.
- Incorrect determination of the numbers of degree of freedom.

Incorrect computations;

The number of applications of chi – square test does not seem to be increasing but the number of misuses of χ^2 has become surprising by large.

A lesson can be drawn from the findings of the Lewis and Burke article.

The chi – square distribution;

For large sample sizes, the sampling (probability) distribution of χ^2 can be closely approximated by a continuous curve known as the chi – square distribution, the probability function of χ^2 distribution is given by;

$$f(x^2) = c(x^2)^{U/2-1} - x^2/2$$

Where,

$$e = 2.71828$$

U = number of degree of freedom

C = a constant depending only on u.

The χ^2 distribution has only one parameter, U, the number of degree of freedom. As in case of t- distribution for each different number of degree of freedom. For very small numbers of degree of freedom, the chi – square distribution is severely skewed to the right, as the number of degree of freedom increase, the curve rapidly becomes more, symmetrical for large values of 'U' the chi – square distribution is closely approximated by the normal curve.

The following diagram gives χ^2 distribution for and degree of freedom

It is clear from the given diagram that as the degree of freedom increase, the curve becomes more and more symmetric.

The chi – square distribution is a probability distribution and the total area under the curve in each chi – squares distribution is like the t- distribution so many distribution is 1.0 of chi – squares are possible that it is not practical to construct a table that illustrates the area under the curve for all possible values of the area. In the appendix on table (for χ^2 only) the area in the table most commonly used in significance tests using the χ^2 distribution are given.

χ^2 Test when the Df exceed so,

$$Z = \sqrt{2\chi^2} - \sqrt{2V - 1}$$

Conditions for applying χ^2 test Nov – 2016;

The following conditions should be satisfied before applying the χ^2 test:

- 1) In the first place N must be reasonably large to ensure the similarity between theoretically correct distribution and our sampling distribution of χ^2 , the chi – square statistic, it is difficult to say what constitute as largeness, but as a general rule χ^2 test sound not be used when N is less then 50 however few the cells.

No theoretical cell frequency should be small when the expected frequency are to small, the value of χ^2 will be over estimated and will

result in too many rejections of the null hypothesis. To avoid making incorrect inference, a general value is followed that expected frequency of less than 5 in one cell of a contingency table is too small to use. When the table contains more than one cell with an expected frequency of less than 5 use 'POOL' the frequencies which are less than 5 with the preceding or succeeding frequency so that the resulting sum is 5 or more. However in doing so, we reduce the number of categories of data and will gain less information from contingency table.

The constraints on the cell frequencies of any should be linear, i.e., they should not involve square and higher powers of the frequencies such as $\sum O = \sum E = N$ frequency so that the resulting sum is 5 or more. However, in doing so, we reduce the number of categories of data and will gain less information from contingency table.

The constraints on the cell frequencies of any should be linear, i.e., they should not involve square and higher powers of the frequencies such as $\sum O = \sum E = N$

SIGN TEST

Meaning

The sign test is a statistical test to compare the sizes of two groups. It is a non parametric or distribution free test which means the test doesn't assume the data comes from a particular distribution. The sign test is an alternative to a one sample t-test or a paired t-test. It can also be used for ordered categorical data.

Assumptions for test also

The data should be from two samples.

The two dependent samples should be paired or matched.

H_0 = No difference in median of the signed differences.

H_1 = median of the signed difference is less than zero.

Calculation (or) Procedures for testing a sign test:

Step 1 : Subtract set 2 from set 1 and put the result in 3rd column.

S.No	Set 1	Set 2	Set 1 – Set 2	Sign
1	443	57	386	+
2	436	587	151	-

Step -2 : Add a fourth column indicating the sign of the number in column 3

Step -3 : Count the number of positives and negatives.

Step -4 : Find the P-value using a binomial distribution.

The critical value of a two sided alternative at $\alpha = 0.05$ can be conveniently found by the expression.

$$K = \frac{(n-1)}{2} - (0.98)\sqrt{n}$$

H_0 is rejected if $S \leq k$ for the sign test.

Types of sign test:

The sign test can be of two types:

One sample sign test

Paired sample sign test

One sample sign test

In a one sample sign test we test the null hypothesis M_0 . Against an appropriate alternative on the basis on a random sample of size n we replace each sample value greater than, M_0 with a plus sign and each sample value less than M_0 with minus sign and discard sample value. Exactly what equal to (M_0). we then test the null hypothesis that these plus and minus signs are values of a random variable having a binomial distribution with $P=1/2$

2. Paired sample sign test;

The sign test has very important applications in problems involving paired data such as data relating to the collection of an account policy, responses of mother and daughter towards ideal family size, etc..., in these problems each pair of sample value can be replaced with a plus sign if the first value is greater then the second, a minus sign if the first value is smaller then the second or be discarded, if the two values are equal then we proceed in the same manners as in one sample sign test.

THE MANN – WHITNEY U TEST;

The sign test for comparing two population of distribution ignore the actual magnitude will go to those of the other sample.

The test of the null hypothesis that the two samples come from identical populations may either be based on R_1 the sum of the ranks of the values of first sample or on R_2 the sum of the ranks of values of second samples, it may be noted that in practice it does not matter which sample we call sample1 and which we call sample2.

If the sample sizes are n_1 and n_2 the sum of R_1 and R_2 is simply the sum of first $n_1 + n_2$ positive integers. Which is known to be $\frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2}$ this formula enables us to find R_2 if we know R_1 and vice versa. When the use of the rank sums was first proposed as a non-parametric alternative of the two samples t-test, the decision was based on R_1 or R_2 , but how the decision is usually based on either of the related statistics.

$$U_1 = (n_1 + n_2) + \frac{(n_1 + n_2 + 1)}{2} - R_1$$

$$U_2 = (n_1 + n_2) + \frac{(n_1 + n_2 + 1)}{2} - R_2$$

Where n_1 and n_2 are the size of the samples and R_1 and R_2 are the ranks sums of the corresponding samples. For small samples if both n_1 and n_2 are less than 10 special tables must be used and if U is smaller than the critical value. He can be related to the standard normal curve by the statistic

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{n_1 n_2 \frac{(n_1 + n_2)}{12}}}$$

In using this statistic, it is unimportant whether the larger or smaller value obtained from the formula is used. The values for Z will be numerically equal, but opposite in sign, note that tied observations are gain given the mean of the common ranks.

THE KRUSKAL – WALL OR H – TEST;

If several independent samples are involved, analysis of variance is the usual procedure failure to meet the assumptions needed for analysis of variance makes its value doubtful. An

alternative technique was developed called the Kruskal Wallis one way analysis of variances or the null hypothesis that K independent random samples come from identical populations against the alternative hypothesis that the means of these samples are not all equal.

As it is done in the Mann Whitney V – test all data we ranked as if they were in one sample, from lowest to highest. The rank sums of each sample are calculated the H- statistic calculated from the formula,

$$H = \frac{12}{N(N+1)} \left[\frac{R_1}{n_1} + \frac{R_2}{n_2} + \dots + \frac{R_k}{n_k} \right] - 3(n+1)$$

When n_1, n_2, \dots, n_k are the number in each of K samples,

$N = n_1 + n_2 + \dots + n_k$ and $R_1 + R_2 + \dots + R_k$ are the rank sums of each sample, if there are ties, the usual procedure is followed, but H is fairly sensitive to this, so if there are very many of them a correction should be made. The effect of the correction is to increase slightly the value of H, so its use is not imperative. For small samples H, is approximately distributed as chi – square with $K-1$ degrees of freedom and table b given in the appendix can be used.

If the null hypothesis is true and each sample has at least five observations the sampling distribution of H can be approximated closely with $K-1$ degrees of freedom with a chi - square distribution with $K-1$ degrees of freedom. Consequently we can reject the null- hypothesis that $\mu_1 = \mu_2 = \dots = \mu_k$ and accept the alternative that the μ_s are not all equal at $K-1$ level of significance α for $H < \chi^2_{\alpha, K-1}$ degrees of freedom. If any sample has less than 5 used, and the test must be based on special tables.

University Question: Nov-2018

What are the advantages of non-Parametric tests?

Advantages of non – Parametric tests;

The major advantages of non- parametric tests are given below;

- They can be applied to a large number of situations.
- They can be more easily understood intuitively.
- They can be used with smaller sample sizes.
- They can be used with more types of data.

- They need fewer or less stringent assumptions about the nature of the population distributions.
- They are generally more robust and not often seriously affected by the extreme values in data such as outliers.
- They have in many cases, a high level of asymptotic relative efficiency compared to the classical parametric tests.
- The introduction of jack knife boot stop and other resembling techniques has increased their range of applicability.
- They provide a number of supplemental or alternative tests and techniques to currently existing parametric tests.

1) 200 digit are a choose at random from the set of table, the frequently of a digit are as follows

Observed frequency:

18	19	23	21	16	25	22
20	21	15				
0	1	2	3	4	5	6
7	8	9				

Use Chi Square test to assess the correctness of the hypothesis that the digit were distributed in equal number. The table from which there were use it.

$$\chi^2_{n-1} = 10-1 = 9$$

$$T.V_{0.05} = 16.919$$

$$T. Value of 0.01 = 21.69$$

Hypothesis

Let us take the hypothesis that the digit were in equal number

$$\Sigma.F = \Sigma f / N = 200 / 10 = 20$$

O	E	O-E	(O-E) ²	(O - E) ² / E
18	20	-2	4	0.2

19	20	-1	1	0.05
23	20	3	9	0.45
21	20	1	1	0.05
16	20	-4	16	0.8
25	20	5	25	1.25
22	20	2	4	0.2
20	20	0	0	0
21	20	1	1	0.05
15	20	-5	25	1.25
				4.3

Conclusion:

Since the calculated value is $4.3 < 21.67$ at 1% Level of Significance. Hence our hypothesis is accepted.

2) In an experiment on pea preening obtained the following frequency of seed. 315 round and yellow, 01 wrinkled and yellow 108 round and green 32 wrinkled and green.

According to his theory of heredity the number should be a proposition. 9:3:3:1 is there any evident his down the theory at 5% level of significance.

C/S	R	W	T
Y	315	101	416
G	108	32	140
Total	423	133	556

Hypothesis

Let us take the hypothesis that there is no significant difference between the observed frequency and expected frequency
Expected Frequency

Color/Shape	R	W	Total
Y	312.75	104.25	417
G	104.25	34.75	139
Total	417	139	556

Ratio \rightarrow 9:3:3:1

$$556 \times \frac{9}{16} = 312.72$$

$$556 \times \frac{3}{16} = 104.25$$

$$556 \times \frac{3}{16} = 104.25$$

$$556 \times \frac{1}{16} = 34.75$$

$$D.F = 4 - 1 = 3$$

Table Value $\chi^2_{0.05} = 7.815$ (5% level significance)

$0.01 = 11.345$ (1% level significance)

O	E	O-E	(O-E) ²	$\frac{(O - E)^2}{E}$
315	312.75	2.25	5.0625	0.016
101	104.25	-3.25	10.5625	0.102
108	104.25	3.75	14.0625	0.135
32	34.75	-2.75	7.5625	0.218
			$\Sigma(O - E)^2 / E$	0.471

$$\Sigma(O - E)^2 / E = 0.471$$

Conclusion:

Since the calculated value of 0.471 < 11.345 at 1% level of significance. Hence our hypothesis is accepted.

3) A movie producer is bringing out a new movie in order to make out his advertising campaign he was determined. Where the movie will appear most the particular age were he appeals equally to all age group. The producer takes the random sample from personal attaining. The preview new movie and obtained the following result

Particular	Age Group				
	Under 20	20-39	40-59	Above 60	Total
Liked the Movie	146	78	48	28	300
Disliked the movie	54	22	42	22	140
In different	20	10	10	20	60
	220	110	100	70	500

What difference will you drop from the data?

$$D.F = n - 1 = 6 \quad T.V = 0.05 = 12.592$$

$$D.F = (c-1)(r-1)$$

$$(4-1)(3-1) = (3 \times 2) = 6 \quad T.V \text{ of } \chi^2$$

Expected Frequency

$$E = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

$$\frac{300 \times 220}{500} = 132$$

$$\frac{140 \times 220}{500} = 61.6$$

$$\frac{60 \times 220}{500} = 26.4$$

$$\frac{300 \times 110}{500} = 66$$

$$\frac{140 \times 110}{500} = 30.8$$

$$\frac{60 \times 110}{500} = 13.2$$

$$\frac{300 \times 100}{500} = 60$$

$$\frac{140 \times 100}{500} = 28$$

$$\frac{60 \times 100}{500} = 12$$

$$\frac{300 \times 70}{500} = 42$$

$$\frac{140 \times 70}{500} = 19.6$$

$$\frac{60 \times 70}{500} = 8.4$$

5) Hypothesis

Let us take the hypothesis that the movie appeals equally to all age groups.

$$E = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

O	E	O-E	(O-E) ²	$\frac{(O-E)^2}{E}$
146	132	14	196	1.48
78	66	12	144	2.18
48	60	-12	144	2.4
28	42	-14	196	4.67
54	61.6	-7.6	57.76	0.94
22	30.8	-8.8	77.44	2.51
42	28	14	196	7
22	19.6	2.4	5.76	0.29
20	26.4	-6.4	40.96	1.55
10	13.2	-3.2	10.24	0.77
10	12	-2	4	0.33
20	8.4	11.6	134.56	16.02
			$\sum(O-E)^2/E$	40.14

Expected Frequency

	132	66	60	42	300
	61.6	30.8	28	19.6	140
	26.4	13.2	12	8.4	60
Total	220	110	100	70	500

Conclusion

Since the calculated value is $40.14 < 12.592$ at 5% level of significance. Our hypothesis is rejected. Hence the movie appeals not equal to all age group.

4) 1000 students at college level are greater according to their economic condition and IQ (Intelligent Question) use chi square to find out where there is only association between the Economic conditions.

Economic Condition	High	Medium	Low	Total
Rich	160	300	140	600
Poor	140	100	160	400
	300	400	300	1000

$$\frac{600 \times 300}{1000} = 180$$

$$\frac{400 \times 300}{1000} = 120$$

$$\frac{600 \times 400}{1000} = 240$$

$$\frac{600 \times 300}{1000} = 180$$

$$\frac{400 \times 400}{1000} = 160$$

$$\frac{400 \times 300}{1000} = 120$$

Expected Frequency

	High	Medium	Low	Total
Rich	180	240	180	600
Poor	120	160	120	400
	300	400	300	1000

$$E = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

O	E	O-E	(O-E) ²	$\frac{(O - E)^2}{E}$
---	---	-----	--------------------	-----------------------

160	180	-20	400	2.22
300	240	60	3600	15
140	180	-40	1600	8.89
140	120	20	400	3.33
100	160	-60	3600	23
60	120	40	1600	13.33
			$\sum(O - E)^2 / E$	65.77

Yates's Correction

- 1) In an experiment an immunization of cat from tuberculosis. Following results were obtained.

Particulars	Affected	Not Affected	Total
Inoculated	12	28	40
Not Inoculated	13	7	20

Enumerate the effect of vaccine in controlling the incidence of the disease.

Hypothesis

Let us take the hypothesis that the vaccine is not effective in controlling the tuberculosis disease.

By applying yate's correction the self frequency is adjusted on the table is rewritten as follows.

Yate's Correction self observed frequency:

Particulars	Affected	Not Affected	Total
Inoculated	12.5	27.5	40
Not Inoculated	12.5	7.5	20
	25	35	60

$$D.F = \text{Row} - 1 \times \text{Column} - 1 \\ = (2 - 1) (2 - 1) = 1$$

Table Value of 1

$$\chi^2_{0.05} = 3.84$$

$$\frac{40 \times 25}{60} = 16.66 \quad \frac{40 \times 35}{60} = 23.33 \quad \frac{20 \times 25}{60} = 8.33$$

$$\frac{20 \times 35}{60} = 11.66$$

O	E	O-E	(O-E) ²	$\frac{(O-E)^2}{E}$
12.5	16.66	-4.17	17.38	1.043
27.5	23.33	4.17	17.38	0.745
12.5	8.33	4.17	17.38	2.086
7.5	11.66	-4.17	17.38	1.490
			$\sum(O-E)^2/E$	5.364

Conclusion

Since the calculated value $\chi^2 = 5.364$ is < 3.84 (table value of 5%). So our hypothesis is rejected.

Non-Parametric Test

1. χ^2 = Chi - Square Test
2. Sign Test (K)
3. KruskalVall's Test
4. Man Whitney Test

2. Sign Test (K)

1) A typing school claimed that the six weeks trained the student type average 60 words per minutes. The random sample of 15 graduates given the types test. The median number of words, per minutes typed by each of the student given below test the hypothesis, that the medium typing speed of graduate is at least 60 words per minute.

Name of the students

A	B	C	D	E	F
G	H				
81	76	53	71	66	59
88	73				
I	J	K	L	M	N
O					
80	66	58	70	60	56
55					

60 Words per minutes

Hypothesis: Let us take the hypothesis that the medium typing speed of graduates is at least 60 words per minutes.

$$K = \frac{n-1}{2} - (0.98) \sqrt{n}$$

Candidates	60 words per minutes	x - 60
A	81	21
B	76	16
C	53	-7
D	71	11
E	66	6
F	59	-1
G	88	28
H	73	13
I	80	20
J	66	4
K	58	-2
L	70	10
M	60	0
N	56	-4
O	55	-5

$$(+)=9 \quad (-)=5 \quad 0=1 \quad n=9+5=14$$

$$K = \frac{n-1}{2} - (0.98) \sqrt{n} = \frac{14-1}{2} - 0.98 \sqrt{14}$$

$$= 6.5 - 0.98 \times 3.741 = 6.5 - 3.666$$

$$K = 2.84$$

K' is less S = accepted

K is greater than S = Rejected

$$S = 5$$

2.84 < 5 so the hypothecation is accepted.

Conclusion

Hence that the median typing speed of graduates is not at least 60 words per minute.

2) As the sign test to see whether there is difference between the number of days utilize collection of an amount received before and after a new collections policy use 0.05 significance level.

Before	30	28	34	35	40
42	33	38			
After	32	29	33	32	37
43	40	41			
Before	34	45	28	27	25
41	36				
After	37	44	27	33	30
38	36				

Before	After	B - A
30	32	-2
28	29	-1
34	33	1
35	32	3
40	37	3
42	43	-1
33	40	-7
38	41	-3
34	37	-3
45	44	1
28	27	1
27	33	-6
25	30	-5
41	38	3
36	36	0

$$+ = 6 \qquad - = 8 \qquad 0 = 1$$

$$n = 14$$

$$K = \frac{n-1}{2} - (0.98) \sqrt{n} = \frac{14-1}{2} - 0.98 \sqrt{14}$$

$$= 6.5 - 0.98 \times 3.741 = 6.5 - 3.666$$

$$K = 2.84$$

$$S = 8$$

$2.84 < 8$. So it is accepted.

3. KRUSKAL WALLIS (or) H-Test

The company trains are random the ascertain to group though a certain initial instruction procedure by 3 different method. At the end of the instructing period they are tested for infections performance quality. The following are they scores.

Method A	Method B	Method C
80	82	93
83	84	65
79	60	77
85	72	78
90	86	88
68	67	-
-	91	-

Use the H-Test to determine at the 0.05 level of significance. Whether the 3 methods are equally effective.

$$H = \frac{12}{N(N-1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} + \frac{R_4^2}{n_4} \right) - 3(NH)$$

H-Test

Solution:

60	1
65	2
67	3
68	4
72	5
77	6
78	7
79	8
80	9
82	10
83	11
84	12
85	13
86	14
88	15

A	R ₁	B	R ₂	C	R ₃
80	9	82	10	93	18
83	11	84	12	65	2
79	8	60	1	77	6
85	13	72	5	78	7
90	16	86	14	88	15
68	4	67	3	-	-
-		91	17	-	-
	61		62		48

90	16
91	17
93	18

$$\begin{aligned}
 H &= \frac{12}{N(N-1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} + \frac{R_4^2}{n_4} \right) - 3(NH) \\
 &= \frac{12}{18(18+1)} \left(\frac{61^2}{6} + \frac{62^2}{7} + \frac{48^2}{5} \right) - 3(18 + 1) \\
 &= \frac{12}{18(19)} \left(\frac{3721}{6} + \frac{3844}{7} + \frac{2304}{5} \right) - 3(19) \\
 &= \frac{12}{342} (620.17 + 549.14 + 460.8) - 57 \\
 &= 0.035 (1630.11) - 57 \\
 &= 57.197 - 57 \\
 H &= 0.197
 \end{aligned}$$

Hypothesis: Let us take the hypothesis that the course of 3 methods are equally effective in testing the performance quality.

$$v = c-1 = 3 - 1 = 2$$

Table value 5% level of significance H Test 0.05 = 5.991

Conclusion

The calculated value of H Test (0.197 < 5.991) at 5% level of significance, hence the hypothesis is accepted and we concluded that the 3 methods are not equally effective in recording information of performance quality.

4. A Rank sum Test (or) The Mann Whitney 'U' Test

Appliance for a passion are centre views by 3 administrators and retail and a scale of 5 has to suitability for the portion of each appliance is given a suitability score which is the sum of 3 members. All through college education is not a requirement for a passion personal director. Felled their it might have sum bring on suitability for a passion. Ratios made that rating on the basis of individual interviews and 12 of the college.

You, the Mann Whitney U-Test the determined whether there was difference in the course of 2 groups use the 0.05 level of 2 groups use the 0.05 level of significance. Group A – education background in and 2 years of college. While Group B has completed at least 2 years in the college.

Suitability Scores.

Group A	11	11	9	4	8	6	12	11	19	10	11
Group B	8	9	13	14	11	10	12	14	13	9	10

$$U = \frac{U_1 + U_2}{2}$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$Z = \frac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n_1 + n_2 - 1) / 12}}$$

Rank

4, 6, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 12, 12, 13, 13, 14, 14

1	8.5	16
2	12	19.5
3	12	19.5
5	12	21.5
5	16	21.5
5	16	23.5
8.5	16	
8.5	16	
8.5		

Group A	R_1	Group B	R_2
7	3	8	5
11	16	9	8.5
9	8.5	13	21.5
4	1	14	23.5
8	5	11	16
6	2	10	12
12	19.5	12	19.5
11	16	14	23.5
9	8.5	13	21.5

10	12	9	8.5
11	16	10	12
11	16	8	5
	123.5		176.5

$$U = \frac{U_1 + U_2}{2}$$

$$\begin{aligned} U_1 &= n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \\ &= 12 * 12 + 12 \frac{(12+1)}{2} - 123.5 \\ &= 144 + 12 \frac{(13)}{2} - 123.5 \\ &= 144 + 78 - 123.5 \\ &= 222 - 123.5 \end{aligned}$$

$$U_1 = 98.5$$

$$\begin{aligned} U_2 &= n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \\ &= 12 * 12 + 12 \frac{(12+1)}{2} - 176.5 \\ &= 144 + 12 \frac{(13)}{2} - 176.5 \\ &= 144 + 78 - 176.5 \end{aligned}$$

$$U_2 = 45.5$$

$$U = \frac{U_1 + U_2}{2} = \frac{98.5 + 45.5}{2}$$

$$U = 72$$

$$\begin{aligned} Z &= \frac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n_1 + n_2 - 1) / 12}} = \frac{72 - 12 * 12 / 2}{\sqrt{12 * 12 (12 + 12 - 1) / 12}} \\ &= \frac{-36}{\sqrt{144(24-1) / 12}} = \frac{-36}{276} \end{aligned}$$

$$Z = -2.167$$

$$v = n_1 + n_2 - 1 = 12 + 12 - 1 = 23$$

Table value @5% level of significance 1.96

$$Z_{0.05} = 1.96$$

Conclusion

Calculate the value of Z (-2.19 < 1.96) at 5% level of significance. Hence the hypothesis is accepted. Hence there is no significant difference between the scores in 2 group.

Chapter – IV

Statistical Quality Control

The term 'quality' in statistical quality control is usually related to some measurement made on the items produced, a good quality item having one which conforms to standards specified for measurement. Quality does not always imply the highest standards of manufacture, for the standard required is often deliberately below the highest possible. It is almost always consistency in quality standards which represents the most desirable situation rather than the absolute standard which is maintained.

The needs for quality control arises because of the fact that even after the quality standards have been specified some variation in quality is unavoidable. For example, a machine is producing 100,000 bolts per day of 2" length. It is very unlikely that all the bolts are exactly 2" in length. If the measuring instrument is sufficiently precise we can detect some screws which are slightly less than 2" and some which are slightly more than 2". This leads to a search in the possible causes of variation in the product. The variation of a quality characteristics can be divided under two heads:-

- i) Chance variation, i.e., variation which result from many minor cause that behave in a random manner. This type of variation is permissible, and indeed inevitable, in manufacturing. There is no way in which it can completely be eliminated-when the variability present in a production process is confined to change variation, the process is said to be in a state of statistical control.
- ii) Assignable variation, i.e., those variations that may be attributed to special non-random causes. Such variations can be the result of the several factor such as a changes in the raw material, a new operation, improper machine setting, broken or worn parts, mechanical faults in plant, etc.

Out of these two types of variation nothing can be done about the former type. However, assignable variation can be detected and corrected. The value of quality control lies in

the fact that assignable variations in a covered before the product becomes defective.

There are two different ways of controlling the quality of a product:

- (i) Through 100% inspection, i.e., by inspecting each and every item that is produced and;
- (ii) Through sampling technique or the use of statistical quality control.

The system of 100% inspection is not very satisfactory because of the following reasons:

- (i) It is too expensive.
- (ii) It is not always reliable because it becomes too much a routine for the persons inspecting each every item and defective pieces may also be labeled 'satisfactory'. Defective pieces may also be passed at times when distraction occurs. For example, even when an inspector is trying to perform his task conscientiously, when somebody talks to him or someone else happens to attract his attention he may at times pass family pieces.
- (iii) The inspection is made at the end manufacturing cycle, and hence provides few controls over the manufacturing process.

Thus we find that even 100% inspection is not infallible. In an effort to find out a more economical and yet practical procedure greater and greater use is being made of the tool of statistical quality control. Statistical quality control is simply a statistical method for determining the extent to which quality goals are being met without necessarily checking every item produced normal expectations. The statistical control of quality makes use of the theory of sampling and tests of significance.

A control chart is essentially a graphic device for presenting data so as to directly reveal the frequency and extent of variations from established standards or goals. Control charts are simple to construct and easy to interpret and they tell the manager at a glance whether

or not the process is in control, i.e., within the tolerance limits. A control chart consists of three horizontal lines:

- i) A control line to indicate the desired standard or level of the process;
- ii) Upper control limit ; and
- iii) Lower control limit.

From time to a sample is taken and the data are plotted on the graph. So long as the sample points fall within the upper and lower control limits there is nothing to worry as in such a case the variation between the samples is attributed to chance or unknown causes.

Types of control charts

Broadly speaking, control charts can be divided under two heads:

- i) Control charts of variables, and
- ii) Control charts of attributes.

Variables are those quality characteristics of a product which are measurable and can be expressed in specific – units of measurement such as diameter of radio knobs which can be measured and expressed in centimeters, tensile strength of cement which can be expressed in specific measures per square inch of space, etc. attributes, on the other hand, are those product characteristics which are not amenable to measurement. Such characteristics can only be identified by their presence or absence from the product. For example , we may say that plastic is cracked or not cracked, whether the bottles that have been manufactured contain holes or not. Attributes may be judged either by the proportion of units that are defective or by the number of defects per unit. Thus the data resulting from inspection of a quality characteristics may take any one of the following forms:

- i) A record of the actual measurements of the quality characteristics for individual articles or specimens.
- ii) A record of number of articles or specimens inspected and of the number found defective.

- iii) A record of the number of defects that are found in a sample. The number of defects per sample may be very large compared to the average number of defects per sample.

For purpose of control, data of the first (i) form listed above may be summarized by taking two statistical measures, the average (\bar{X}) and the standard deviation (σ), or the average (\bar{X}) and the range (R). Data of the second form (ii) can be summarized in terms of fraction defective (p), and data of type (iii) can be summarized in terms of number of defects per unit (c).

Setting up a control procedure

In establishing basic procedure for the operation of a quality control program, the manufacturer must take the following preliminary steps:

1. Select the quality characteristics that are to be controlled (including the limits of variation).
2. Analyze the production process to determine the kind and location of probable cause of irregularities.
3. Determine how the inspection data are to be collected and recorded, and how they are to be sub – divided.
4. Choose the statistical measures that are to be used in the charts.

Depending on the type of inspection data available, any one of the following types of control charts may be used.

1. *Control charts \bar{X} and σ and \bar{X} and R .* Such charts are used when measured values of the quality characteristics are at hand.
2. *Control chart for \bar{X} alone.* Control chart for \bar{X} alone is used where experience with control charts for \bar{X} and R , or \bar{X} and σ has demonstrated that instances of lack of control are almost always associated with causes that affect \bar{X} rather than σ or R .
3. *Control chart for σ or R alone.* Control chart for R or σ is used alone where technical reasons render control of \bar{X}

unimportant or where control for \bar{X} is known to be unjustifiably expensive.

4. *Control chart for c.* This chart is used in such situations where in the inspection consists of determining the number of defects c in a sample. Such is the case, for example, in the examination of finished textiles, plywood sheets etc.
5. *Control chart for p or pn .* Chart for p or pn is used when the records of inspection or testing show merely the number of articles inspected and the number found defective.

Statistical Decision Theory

First, decision is needed in a problem situation where two or more alternative courses of action are available, and where only one of these actions can be taken. Obviously, if there is only one course of action available, no decision is required since that action must be taken in order to solve the problem.

The decision procedure involves selecting among the alternatives a single course of action that can be actually carried out. If such a course of action is selected that cannot be carried out in the existing situations and circumstances, it will tant amount to waste of time and resources. Quire often the objective of decision is to select an act which will accomplish some predestinated purpose. The decision taken may be regarded as satisfactory or not depending upon whether it has helped in the attainment of that objective.

Secondly, in all decision problems "uncertainty" is found to be a common element. When the outcome of some action is not known in advance, the outcome is said to be uncertain. When there are many possible outcomes of an event (also called states of nature) one cannot predict with certainty what will happen – it is only in terms of probabilities we may be able to talk.

Lastly, a number of consequences result from each action under different conditions, the conditions being various states of nature in general if there are m possible actions and n admissible states of nature, the consequences will be $n \times m$ in number In practical situations particularly in business and economic problems consequences can be expressed in terms of money and utility.

The consequences may be evaluated in several ways such as

:

- i) In terms of profit
- ii) In terms of cost, and
- iii) In terms of opportunity loss.

In most decision problems the expected monetary value is used as a decision criterion. When consequences are evaluated in terms of profit, they are called payoffs. A payoff table is prepared and it shows the relation between all possible states of nature, all possible actions and the values associated with the consequences. A specimen of payoff table is given below:

Optimal decisions. After we have been able to prepare a payoff or loss table it is easy to arrive at a decision and to select a best course of action. It would be desirable to choose that action which maximizes payoffs or minimizes opportunity losses for all states of nature. If such action exists, the action is called uniformly best action. Unfortunately, such uniformly best action exists very seldom, because a particular action may be best for some states of nature and the same action be worst for the remaining states of nature. For instance consider illustration 1 where it is clear that the action a_3 is best when the state of nature is θ_3 . On the other hand, the same action is worst when the state of nature is θ_1 , and none of the actions is found to be uniformly best. Similar situation exists in illustration 2. Hence, the decision-maker needs some criterion or principle for making a choice amongst alternative actions. We shall mention here some important principles or criteria to select an optimal action. An action is said to be optimal if its payoff is best (as large as possible) according to the criterion or principle under consideration.

1. **The maximin principle.** Of all the principles the simplest is the maximum principle for determining an optimal action when consequences are given in term of profits. According to this principle the decision-maker first observes the minimum payoffs over the various possible states of nature. Then he selects that action for which the minimum payoff is maximum. This principle places a value on each action according to the

worst that can happen with that action in a sense, decision-maker expects worst and prepares for it.

2. **The minimax principle.** This principle is used only when consequence are given in the form of opportunity losses which is same as loss table. The decision-maker first observes the maximum opportunity loss over all the various states of nature. He then selects that action for which the maximum opportunity loss is minimum. This principle gives the greatest possible protection against the largest loss.
3. **The Baye's principle.** The Bayes principle for the selection of an optimal action derives its name from the 18th century philosopher Thomas Bayes, who first suggested and investigated the notion of "inverse probability" or "subjective probability". The disciples of Bayesian School regard probability as a measure of the degree of personal belief and they prefer to be named as subjectivists. On the other hand, the disciples of non-Bayesian school consider probability as a long run relative frequency and they prefer to be labeled as objectivists. One major advantage of the Bayesian approach is that the decision-maker selects a course of action on a rational basis since he uses subjective evaluation of probability based on experience, past performance, judgment, etc.

After determining the prior distribution, the Bayes principle is to be used phase wise. The three phases in order of their occurrence are: (i) prior analysis, (ii) preposterior analysis, and (iii) posterior analysis. A brief description of each of these is given below:

- (i) **Prior Analysis.** Once the relevant prior distribution of various states of nature is found out, the decision-maker needs to compute the expected payoff (abbreviated EP) or expected opportunity loss (EOL) for each action.
- (ii) **Preposterior Analysis.** After making prior analysis the decision-maker must decide either to collect additional information regarding the states of nature or to take the action as suggested by the

prior analysis. Prior distribution is not always a perfect predictor regarding the states of nature. This is more so in business decision problems. However, if some how, the decision-maker finds a perfect predictor, he would prefer actions based on perfect predictor for it would enable him to maximize his profits or minimize his losses.

- (iii) **Posterior Analysis.** If it is decided to gather further information regarding the states of nature, the information to be gathered by conducting survey or by performing an experiment or by some other means the posterior analysis is used only after gathering the information. In posterior analysis relevant information is combined with prior information in order to make the degree of belief regarding the states of nature more stronger. In other words the decision-maker has to revise the prior distribution as a result of which he gets new distribution of states of nature, called the posterior distribution. If the additional information is collected by conducting a survey the outcome of he survey would convert the prior distribution into the posterior distribution. The decision-maker will ultimately be interested in the expected posterior payoffs.

Statistical Decision Theory

Cost = 3; Selling Price = 5; Profit = 2.

A Baker producer's a certain type of special party at a cost of Rs.3 and sells it at the price of Rs.5. This pastry is produce over the week and sold during the week. Such pastries being produce but not sold during the week time is totally spoiled and have to be thrown away. According to cost experience, the weekly demand for these pastries never less than 78 (or) greater than 80. You are required to formulate action space pay-off table and pay-loss table.

Solve:

- 1) There are 3 courses of action that is $A_1 = 78$; $A_2 = 79$; $A_3 = 80$.
Production can be made.
- 2) Thus the action space are a is equal to $(a_1 + a_2 + a_3)$
- 3) State the nature weekly demand for pastries there are 3 possible states in nature.

$Q_1 =$ Demand is 78%

$Q_2 =$ Demand is 79%

$Q_3 =$ Demand is 80%

Pay - off Table

State of Nature	Supply	Q_1 78	Q_2 79	Q_3 80	Maximum payoff
	78	(78x3)- (78x3) 156	$79-78=1$ (1x3=3) (156-3) 153	$80-78=2$ $2 \times 3=6$ (156-6) 150	156
79	$78-79=1$ $1 \times 2=2$ (158- $2=156$)	$(79 \times 5)-$ (79x3) 158	$80-79=1$ $1 \times 3=3$ (158-3) 155	158	
80	$78-80$ $2 \times 2=4$ $160-$ $4=156$ 156	$79-80=1$ $1 \times 2=2$ (160-2) 158	$(80 \times 5)-$ (80x3) 160	160	

Pay-loss Table

State of Nature	Supply/Action	a_1 78	a_2 79	a_3 80
	78	(156-156) 0	(156-153) 3	(156-150) 6
79	(158-156) 2	(158-158) 0	(158-155) 3	
80	(160-156) 4	(160-158) 2	(160-160) 0	

The bakery take a choose the 3rd action, which as a maximum pay-off Rs.160.

2) Under an employment promotional program, it is proposed to allow sale of newspapers on the buses during off peak hours. The vender can purchase a newspaper at a special concession rate of 25 paise per copy against the selling price of 40 paise. Any unsold copy are however a died loss vendor has estimated the following probability distribution of the no. of copies demanded.

No.of copies	15	16	17	18	19	20
Probability	0.04	0.19	0.33	0.26	0.11	0.07

How many copies sold is order so that he expected profit will be maximum.

Supply	15	16	17	18	19	20
Demand 15	(15x0.15) 2.25	(1x0.25) 2.00	(2.25-0.75) 1.75	(2.25-1.00) 1.50	(2.25-1.00) 1.25	(2.25-1.25) 1.00
16	2.25	(16x0.15) 2.40	(2.40-0.25) 2.15	(2.40-0.50) 1.90	(2.40-0.75) 1.65	(2.40-1.01) 1.40
17	2.25	2.40	(17x0.15) 2.55	(2.55-0.25) 1.90	2.55-0.5) 1.65	(2.55-0.75) 1.8
18	2.25	2.40	2.15	(18x0.15) 2.70	(2.70-0.25) 1.65	(2.70-0.50) 2.2
19	2.25	2.40	2.15	1.90	(19x0.15) 2.85	(2.85-0.25) 2.6
20	2.25	2.40	2.55	2.70	2.85	(20x0.15) 3.00

Supply	15	16	17	18	19	20
Demand (0.04) 15	(2.25x0.04) 0.09	(2.00x0.04) 0.08	0.07	0.06	0.05	0.04
(0.19) 16	0.43	0.46	0.41	0.36	0.31	0.27
(0.33) 17	0.74	0.79	0.84	0.76	0.68	0.59
(0.26) 18	0.59	0.62	0.66	0.70	0.64	0.57
(0.11) 19	0.25	0.26	0.28	0.30	0.31	0.29
(0.07) 20	0.16	0.17	0.18	0.19	0.20	0.21
Total	2.26	2.38	2.44	2.37	2.19	1.97

Vendor can purchase 17 number of copies for the maximum profit.

3) A provided of a food stall has introduced the new item of food delicacy the calls "whim". He has calculated that the cost of manufacture is Rs.1 per piece and that because of its novelty and quality. It would be sold for Rs.3 per piece. It is however perishable and any goods unsold at the end of day are a dead loss. He expects the demand to be variable and has drawn up the following probability.

Distribution expressing his estimates.

No. of pieces demanded	:10	11	12	13	14	15
Probability	:	07	10	23	38	12 10

- i) Find an expression for his net profit (or) loss if he manufactures 'm' pieces and n are demanded consider. Separately the 2 cases n is > m, n is < m.
- ii) Assume that the manufacture 12 pieces. Using the result in 1 above find his net profit/loss for each level of demand.
- iii) Using the probability distribution. Calculate the expected net profit/loss if he manufacture 12 pieces.

- iv) Calculate similarly expected profit/loss for each are the other level of manufacture ($10 < m < 15$)
- v) How many pieces should be manufacture so that his not expected. Profit is maximum.

Solution:

The proporated these not produce more than 15 pieces of "WHIM" (or) less than 10 pieces. His profit determined by the demand (n) and production (m). when the demand is more than the production its profit shall be,

Supply	10	11	12	13	14	15
Demand	(10x2)	(20x1)	(20-2)	(20-3)	(20-4)	(20-5)
10	20	19	18	17	16	15
11	20	(11x2) 22	(22-1) 21	(22-2) 20	(22-3) 19	(22-4) 18
12	20	22	(12x2) 24	(24-1) 23	(24-2) 22	(24-3) 21
13	20	22	24	(13x2) 26	(26-1) 25	(26-2) 24
14	20	22	24	26	(14x2) 28	(28-1) 27
15	20	22	24	26	28	(15x2) 30

Supply	10	11	12	13	14	15
Demand	(20x07)	(19x07)	126	119	112	105
10	140	133				
11	200	220	210	200	190	180
12	460	506	552	529	506	483
13	760	836	912	988	950	912
14	240	264	288	312	336	324
15	200	220	240	260	280	300

Total	2000	2176	2328	2408	2374	2304
-------	------	------	------	------	------	------

The probability distribution expressing his estimate is 13 pieces of profit.

- 4) The management is faced with the problem choose in one of the product for manufacturing. The probability matrices of the market research for the 2 product was as follows.

State of nature acts	Good	Fair	Poor
Product A	0.75	0.15	0.10
Product B	0.60	0.30	0.10

The profit that the management can make for different level of market acceptability of the product are as follows.

State of nature acts	Profit (Rs) if market		
	Good	Fair	Poor
Product A	35000	15000	5000
Product B	50000	20000	(-3000) loss

Calculated expected value of choice and advice the management.

- 5) Calculated the loss table from the following pay. Off table

Pay off table				
Event				
Action	E ₁	E ₂	E ₃	E ₄
	High value	High value	High value	High value
A ₁	400	300	300	100
A ₂	50	300	-150	50
A ₃	400	0	100	0
A ₄	-50	200	0	100
A ₄	0	300	300	0

Suppose the probability of event in this table are
 $B(E_1 = 0.15)$ $B(E_2 = 0.45)$ $B(E_3 = 0.25)$ $B(E_4 = 0.15)$

Calculated the expected pay off and the expected loss of each action

4) Solution:

Expected pay off table				
Product	Good	Fair	Poor	Total
A	$0.75 \times$	$0.15 \times$	0.10×5000	29000
	35000	15000	= 500	
	= 26250	= 2250		
B	$0.60 \times$	$0.30 \times$	$0.10 \times (-$	35700
	50000	20000	3000) = (-	
	= 30000	= 6000	300)	

Since the expected pay-off for product B is more. Product B should be prepared by the management.

5) Solution:

Opportunity loss table				
Action	E_1	E_2	E_3	E_4
A_1	$(400-50)$	$(300-300)$	$(300-(-150))$	$(100-50)$
	350	0	450	50
A_2	$(400-400)$	$(300-0)$	$(300-100)$	$(100-0)$
	0	300	200	100
A_3	$(400-(-50))$	$(300-200)$	$(300-0)$	$(100-100)$
	450	100	300	0
A_4	$(400-0)$	$(300-300)$	$(300-300)$	$(100-0)$
	400	0	0	100

Expected pay – off table					
Action	Event				Total pay-off for every action expected pay-off (4 event)
	E ₁	E ₂	E ₃	E ₄	
A ₁	(50 x 0.15) 7.5	(300 x 0.45) 135	(-150 x 0.25) -37.5	(50 x 0.15) 7.5	112.5
A ₂	(400- 0.15) 60	(10 x 0.45) 0	(100 x 0.25) 25	(0 x 0.15) 0	85
A ₃	(-50 x 0.15) -7.5	(200 x 0.45) 90	(0 x 0.25) 0	(100 x 0.15) 15	97.5
A ₄	(0 x 0.15) 0	(300 x 0.45) 135	(300 x 0.25) 75	(0 x 0.15) 0	210

Expected loss table					
Action	Event				Total pay-off for every action expected pay-off (4 event)
	E ₁	E ₂	E ₃	E ₄	
A ₁	(350 x 0.15) 52.5	(0 x 0.45) 0	(450 x 0.25) 112.5	(50 x 0.15) 7.5	172.5
A ₂	(0x0.15) 0	(300 x 0.45) 135	50	15	200
A ₃	(450 x	(100 x	75	0	187.5

	0.15)	0.45)			
	67.5	45			
A ₄	(400 x	(0 x	0	15	75
	0.15)	0.45)			
	60	0			

- 6) Foot product company is contemplating the introduction of a revolutionarily new product with new packaging to replace the existing product at a higher price (S_1) or a moralized change in a composition of an existing product with a new packaging at a small increasing price S_2 or a change a small change in the composition of the existing product expect the world new with a negligible increase in price (S_2). The three possible states of nature of events or (one) high increase in sales (N_1) to no change in sales (N_2) and decrease in sales (N_3). The marketing department of the company worked out pay off in terms of yearly net profit for each of the strategies for event (expected sales) this is represented in the following table

State of nature strategy	Payoff (in Rs)		
	N_1	N_2	N_3
S_1	700000	300000	150000
S_2	500000	450000	0
S_3	300000	300000	300000

Which strategy should the executive concern choose on the basis of

- 1) Maximum criterion
 - 2) Maximax criterion
 - 3) Minimax regret criterion
 - 4) Laplace criterion
- 7) A group of student rises money each a by selling souvenirs outside the stadium. After a cricket match between teams A

& B. They can by any of the 3 different type of soreness from a supplies team wins the match. A condition payoff table is as under.

Team	Type of souvenirs		
	1	2	3
A wins	₹.1200	₹.800	₹.300
B wins	₹.250	₹.700	₹.1100

- 1) Construct the opportunity loss table
 - 2) Which type of souvenirs which student by if, probability of team A is winning is (0.6) team (0.4)
- 8) A management is faced with the problem of choosing one of three product for manufacturing. The potential demand for each product may turn out to be good, moderate or poor. The probabilities for each of the state of nature were estimated as follow:

Product	Nature of demand		
	Good	Moderate	Poor
X	0.70	0.20	0.10
Y	0.50	0.30	0.20
Z	0.40	0.50	0.10

The estimated profit or loss under the three states may be taken as:

X	30000	20000	10000
Y	60000	30000	20000
Z	40000	10000	-15000 (loss)

Prepare the expected value table and advise the management about the choice of product.

- 9) Pay off of three acts A,B,C and D and the states of nature P,Q,R are given bellow:

States of nature	Pay off (rupees) acts		
	A	B	C
P	-35	120	-100
Q	250	-350	200
R	550	650	700

The probability of the states of nature are 0.5, 0.1, and 0.4 respectively. Tabulate the expected monetary values for the above data and states which can be chosen as the best act?

6) i) Maximin criterion : (Row wise minimum)

	N ₁	N ₂	N ₃
S ₁ - Strategy	= 150000		
S ₂ - Strategy	= 0		
S ₃ - Strategy	= 300000		

When this criterion is adopted that causes of action is selected which maximum the minimum pay-off. The executive choose (strategy S₂ = 0)

ii) Maximax criterion: (row wise maximum pay-off)

Stategy - S ₁	= 700000
Stategy - S ₂	= 500000
Stategy - S ₃	= 300000

In this criterion be select the strategy which gives the maximum pay-off. The executive should choose (strategy S₁ = 700000).

III) Minimax regret criterion:

A)

Opportunity loss-table

Stategy	N ₁	N ₂	N ₃
S ₁	0	150000	150000
S ₂	200000	0	300000
S ₃	400000	150000	0

(column wise maximum – column value)

B) Maximum opportunity loss:

$$S_1 = 150000$$

$$S_2 = 300000$$

$$S_3 = 400000$$

When the criterion adopted 3 steps are necessary.

The executive should choose (strategy (S_1)) for it minimize the maximum opportunity loss.

IV) Lap lace criterion:

$$S_1 = \frac{700000+300000+150000}{3} = 383333$$

$$S_2 = \frac{500000+450000+0}{3} = 316667$$

$$S_3 = \frac{300000+300000+300000}{3} = 300000$$

Since the lap lace criterion is highest for strategy, the executive should be select (strategy (S_1) = 383333).

7) Opportunity loss table

	Types of souvenir			Probability
	I	II	III	
Team (A)	1200 – 1200 0	1200 – 800 400	1200 – 300 900	0.6
Team (B)	1100 – 250 850	1100 – 700 400	1100 – 1100 0	0.4
Total	(0.6 x 0) + (0.4 x 850) 340	(0.6 x 400) + (0.4 x 400) 400	(0.6 x 900) +(0.4 x 0) 540	

Since EOL is minimum for souvenir I, the students should buy souvenir I.

8) Computation of E.V for various acts

Status of nature	Expected pay off (in Rs.10000) for various acts								
	X			Y			Z		
	X ₁	P ₁	X ₁ P ₁	X ₂	P ₂	X ₂ P ₂	X ₃	P ₃	X ₃ P ₃
Good	30	0.7	21	60	0.5	30	40	0.4	16
Moderate	20	0.2	4	30	0.3	9	10	0.5	5
Poor	10	0.1	1	20	0.2	4	-15	0.1	-1.5
Expected monetary value			26			43			19.5

Since the expected value is highest for II course of action, the management I advised to produce product – Y

9) Expected monetary value for each act

Status of nature	Acts			Probability	Expected monetary value		
	A	B	C	(P)	PA	PB	PC
P	-35	1200	-100	0.5	-1.75	6.00	-50.0
Q	250	-350	200	0.1	250	-35.0	20.0
R	550	650	700	0.4	220	260	280
Total					227.5	285	250

Since, the EMV is highest for act B, therefore it is the test out to be between

- 10) A new paper distributor assign probabilities to the demand for a magazine as follows:

Copies demanded : 1 2 3
4

Probability : 0.4 0.3 0.2 0.1

A copy of magazine sells for ₹.7 costs ₹.6 what can be the maximum possible expected monetary value (EMV) if the distributor can return unsold copies for ₹.5 each

Cost = ₹.6 : sold = ₹.7 : profit = ₹.1
: Loss = ₹.1

Supply Demand	1	2	3	4
1	1x1 = 1	=1 - 1 = 0	+1 -2 = -1	-2
2	1	2	1	0
3	1	2	3	2
4	1	2	3	4

Probability

Supply Demand	1	2	3	4
1	0.4	0	-0.4	-0.8
2	0.3	0.6	0.3	0
3	0.2	0.4	0.6	0.4
4	0.1	0.2	0.3	0.4
	1.0	1.2	0.8	0

If the distributor can return unsold copies for ₹.5 each, the maximum possible expected monetary value in the given case is ₹.1.2