Statistical Technique-II

Unit: II

Subject Name: Mathematics-IV

Subject Code: AAS0301A

Dr. Kunti Mishra
NIET, Gr Noida
Department of Mathematics

B Tech 4th Sem

5/20/2023

BRIEF INTRODUCTION OF FACULTY

Dr. Kunti Mishra
Assistant Professor
Department of Mathematics

Qualifications :
M.Sc.(Maths), M. Tech.(Gold Medalist) in Applied and Computational Mathematics, Ph.D

Ph.D. Thesis : Some Investigations in Fractal Theory
Total Number of Research Papers:15
Area of Interests: Fixed Point Theory, Fractals
Teaching Experience: 9 years

Evaluation Scheme

NOIDA INSTITUTE OF ENGINEERING & TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute)

B. TECH (CSE)
EVALUATION SCHEME
SEMESTER-IV

Sl. No.	Subject Codes	Subject Name	Periods			Evaluation Scheme				End Semester		Total	Credit
			L	T	P	CT	TA	TOTAL	PS	TE	PE		
1	AAS0402	Engineering Mathematics-IV	3	1	0	30	20	50		100		150	4
2	AASL0401	Technical Communication	2	1	0	30	20	50		100		150	3
3	ACSE0405	Microprocessor	3	0	0	30	20	50		100		150	3
4	ACSE0403A	Operating Systems	3	0	0	30	20	50		100		150	3
5	ACSE0404	Theory of Automata and Formal Languages	3	0	0	30	20	50		100		150	3
6	ACSE0401	Design and Analysis of Algorithm	3	1	0	30	20	50		100		150	4
7	ACSE0455	Microprocessor Lab	0	0	2				25		25	50	1
8	ACSE0453A	Operating Systems Lab	0	0	2				25		25	50	1
9	ACSE0451	Design and Analysis of Algorithm Lab	0	0	2				25		25	50	1
10	ACSE0459	Mini Project using Open Technology	0	0	2				50			50	1
11	ANC0402 / ANC0401	Environmental Science*/ Cyber Security*(Non Credit)	2	0	0	30	20	50		50		100	0
12		MOOCs** (For B.Tech. Hons. Degree)											
		GRAND TOTAL										**1100**	**24**

****List of MOOCs (Coursera) Based Recommended Courses for Second Year (Semester-IV) B. Tech Students**

S. No.	Subject Code	Course Name	University / Industry Partner Name	No of Hours	Credits
1	AMC0046	Algorithmic Toolbox	University of California San Diego	24	1.5
2	AMC0031	Data Structures	University of California San Diego	25	2

Syllabus

Unit-I (Statistical Techniques-I)

Introduction: Measures of central tendency: Mean, Median, Mode, Moment, Skewness, Kurtosis, Curve Fitting ,Method of least squares, Fitting of straight lines, Fitting of second degree parabola, Exponential curves ,Correlation and Rank correlation, Linear regression, nonlinear regression and multiple linear regression

Unit-II (Statistical Techniques-II)

Testing a Hypothesis, Null hypothesis, Alternative hypothesis, Level of significance, Confidence limits, p-value, Test of significance of difference of means, Z-test, t-test and Chi-square test, F-test, ANOVA: One way and Two way. Statistical Quality Control (SQC), Control Charts, Control Charts for variables (Mean and Range Charts), Control Charts for Variables (p, np and C charts).

Syllabus

Unit III (Probability and Random Variable)

Random Variable: Definition of a Random Variable, Discrete Random Variable, Continuous Random Variable, Probability mass function, Probability Density Function, Distribution functions.

Multiple Random Variables: Joint density and distribution Function, Properties of Joint Distribution function, Marginal density Functions, Conditional Distribution and Density, Statistical Independence, Central Limit Theorem (Proof not expected).

Unit IV (Expectations and Probability Distribution)

Operation on One Random Variable – Expectations: Introduction, Expected Value of a Random Variable, Mean, Variance, Moment Generating Function, Binomial, Poisson, Normal, Exponential distribution.

Syllabus

Unit V (Wavelets and applications and Aptitude-IV)

Wavelet Transform, wavelet series. Basic wavelets (Haar/Shannon/Daubechies), orthogonal wavelets, multi-resolution analysis, reconstruction of wavelets and applications.

Number System, Permutation & Combination, Probability, Function, Data Interpretation, Syllogism.

Branch Wise Application

- Data Analysis
- Artificial intelligence
- Network and Traffic modeling

Course Objective

- The objective of this course is to familiarize the students with statistical techniques. It aims to present the students with standard concepts and tools at an intermediate to superior level that will provide them well towards undertaking a variety of problems in the discipline.

  The students will learn:
- Understand the concept of correlation, moments, skewness and kurtosis and curve fitting.
- Apply the concept of hypothesis testing and statistical quality control to create control charts.
- Remember the concept of probability to evaluate probability distributions.
- Understand the concept of Mathematical Expectations and Probability Distribution.
- Remember the concept of Wavelet Transform and Solve the problems of Number System, Permutation & Combination, Probability, Function, Data Interpretation, Syllogism.

Course Outcome

CO1: Understand the concept of correlation, moments, skewness and kurtosis and curve fitting.

CO2: Apply the concept of hypothesis testing and statistical quality control to create control charts.

CO3: Remember the concept of probability to evaluate probability distributions

CO4: Understand the concept of Mathematical Expectations and Probability Distribution

CO5: Remember the concept of Wavelet Transform and Solve the problems of Number System, Permutation & Combination, Probability, Function, Data Interpretation, Syllogism.

Program Outcome

S.No	Program Outcomes (POs)
PO 1	Engineering Knowledge
PO 2	Problem Analysis
PO 3	Design/Development of Solutions
PO 4	Conduct Investigations of Complex Problems
PO 5	Modern Tool Usage
PO 6	The Engineer & Society
PO 7	Environment and Sustainability
PO 8	Ethics
PO 9	Individual & Team Work
PO 10	Communication
PO 11	Project Management & Finance
PO 12	Lifelong Learning

PSO

PSO	Program Specific Outcomes(PSOs)
PSO1	The ability to identify, analyze real world problems and design their ethical solutions using artificial intelligence, robotics, virtual/augmented reality, data analytics, block chain technology, and cloud computing
PSO2	The ability to design and develop the hardware sensor devices and related interfacing software systems for solving complex engineering problems.
PSO3	The ability to understand inter disciplinary computing techniques and to apply them in the design of advanced computing.
PSO4	The ability to conduct investigation of complex problem with the help of technical, managerial, leadership qualities, and modern engineering tools provided by industry sponsored laboratories.

CO-PO Mapping(CO1)

Sr. No	Course Outcome	PO1	PO 2	PO 3	PO4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12
1	CO1	H	H	H	H	L	L	L	L	L	L	L	M
2	CO2	H	H	H	H	L	L	L	L	L	L	M	M
3	CO3	H	H	H	H	L	L	L	L	L	L	M	M
4	CO4	H	H	H	H	L	L	L	L	L	L	L	M
5	CO5	H	H	H	H	L	L	L	L	L	L	M	M

*L= Low *M= Medium *H= High

CO-PSO Mapping(CO2)

CO	PSO1	PSO2	PSO3	PSO4
CO.1	H	L	M	L
CO.2	L	M	L	M
CO.3	M	M	M	M
CO.4	H	M	M	M
CO.5	H	M	M	M

*L= Low *M= Medium *H= High

Program Educational Objectives(PEOs)

PEO-1: To have an excellent scientific and engineering breadth so as to comprehend, analyze, design and provide sustainable solutions for real-life problems using state-of-the-art technologies.

PEO-2: To have a successful career in industries, to pursue higher studies or to support entrepreneurial endeavors and to face the global challenges.

PEO-3: To have an effective communication skills, professional attitude, ethical values and a desire to learn specific knowledge in emerging trends, technologies for research, innovation and product development and contribution to society.

PEO-4: To have life-long learning for up-skilling and re-skilling for successful professional career as engineer, scientist, entrepreneur and bureaucrat for betterment of society.

Result Analysis

Branch	Semester	Sections	No. of enrolled Students	No. Passed Students	% Passed
CS	IV	A	67	65	92%
IOT	IV	A	49	45	91.83%

End Semester Question Paper Template

Link: [100 Marks Question Paper Template.docx]

Prerequisite and Recap(CO2)

- Knowledge of Mathematics I ,II and III of B. Tech or equivalent.

Brief Introduction about the Subject with Video Links

- In first four modules, we will discuss Statistics.
- In 5^{th} module we will discuss wavelet Transform & aptitude part.

Video Links:

- https://youtu.be/cKakQI_jl3w
- https://youtu.be/bMSYgnrM-zw
- https://youtu.be/mZ-FREAgCyY

Unit Content

- Testing a Hypothesis,
- Null hypothesis
- Alternative hypothesis
- Level of significance
- Confidence limits
- p-value
- Test of significance of difference of means:
- Z-test
- t-test
- Chi-square test
- F-test
- ANOVA: One way and Two way
- Statistical Quality Control (SQC)
- Control Charts
- Control Charts for variables (Mean and Range Charts)
- Control Charts for Variables (p, np and C charts).

Unit Objective(CO2)

- The objective of this course is to familiarize the engineers with Statistical techniques and its applications in real world.
- The concept of hypothesis, ANOVA and quality control by control charts.

Topic objective (CO2)

Test of Hypothesis

- Describe real-life examples to explain the motivation behind hypothesis

Test of Hypothesis (CO2)

Test of Hypothesis:

- Some information about a characteristic of the population is known, we wish to know whether the information can be accepted or rejected, we choose a random sample and obtain information about the characteristic, based on this information.
- we conclude whether the available information of the characteristic of the population can be accepted or rejected. This is called Test of hypothesis.

Symbols for Population & Samples:

	Population parameter/characteristic	Sample Statistic
Symbols	Population Size (N)	Sample Size (n)
	Population mean (μ)	Sample mean ($\bar{x}$)
	Population Standard Deviation (σ) Population variance (σ^2)	Sample Standard Deviation (s) sample variance (s^2)

Test of Hypothesis (CO2)

Procedure of Test of hypothesis

- State Null hypothesis & Alternate hypothesis
- Define level of significance
- Use suitable test statistics
- Establish critical region (Table value)
- Conclusion

Test of Hypothesis (CO2)

Null Hypothesis

- A population is given to us and we wish to have information about a parameter of the Population.
- we start with the assumption that there is no significant difference between the sample statistic and the corresponding population parameter or between two sample statistic.
- This assumption there is no significant difference is called a null hypothesis and is denoted by H_0.

Test of Hypothesis (CO2)

Alternate Hypothesis

A hypothesis that is different from the null hypothesis is called alternate hypothesis and is denoted by H_1.

Single tailed & two tailed test

Consider the following example-

let the null hypothesis is define as-

H_0: The population has an assume value of

mean μ_0 i.e. $\mu = \mu_0$.

Then alternate hypothesis can be define as any

of the following-

Test of Hypothesis (CO2)

I. $H_1: \mu \neq \mu_0$

II. $H_1: \mu > \mu_0$

III. $H_1: \mu < \mu_0$

Depending upon the problem we use single tail or two tailed test to estimate the significance of result. In two tailed test the areas of the both the tails of the curve representing the sampling distribution are taken into account whereas in single tail test only the area on left or right of is taken into account.

Test of Hypothesis (CO2)

Level of significance

The probability level below which we reject the hypothesis is known as level of significance. It is denoted by α and generally specified before sample are drawn.

Generally we take $\alpha = 5\%$ or 1%.

Critical value

The value of test statistic which separates the critical region (or rejection) and acceptance region is called the critical value and it depends upon

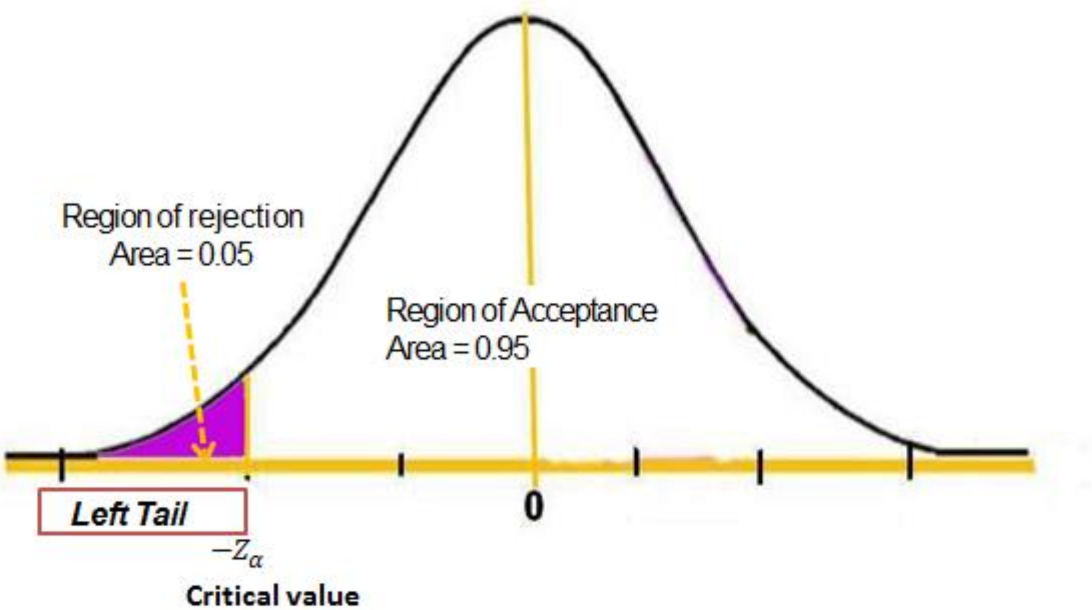
- Level of significance
- Alternate hypothesis (whether it is two tailed or single tailed)

Test of Hypothesis (CO2)

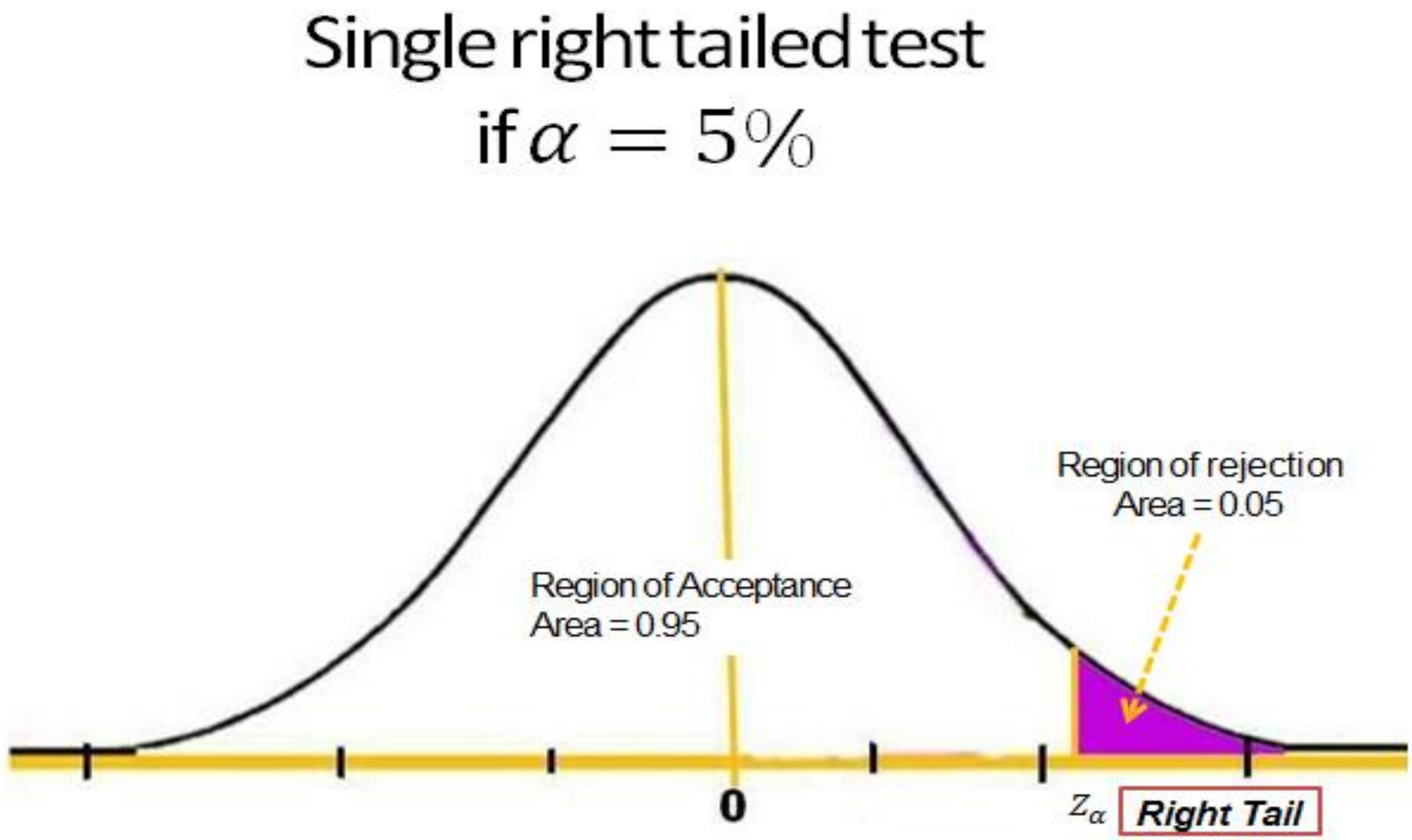
Critical Region or rejection region

Critical region is the region in which a sample value is rejected.

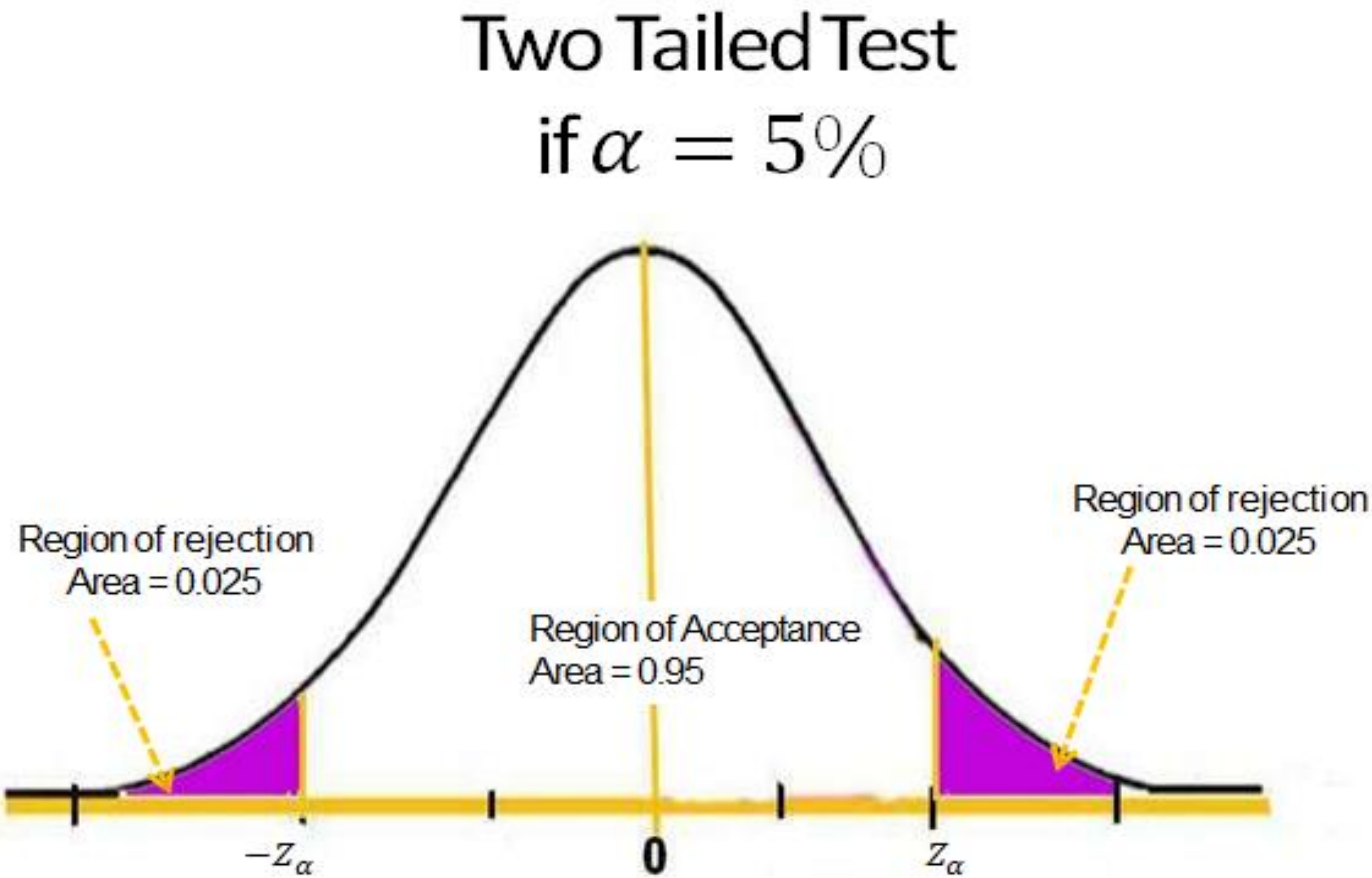
Single left tailed test $\alpha = 5\%$:



Test of Hypothesis (CO2)



Test of Hypothesis (CO2)



Test of Hypothesis (CO2)

Conclusion

If absolute calculated value of test statistic is less than or equal to its critical value(table value), we accept the null hypothesis H_0 otherwise reject it.

Test of hypothesis for large samples and Small samples

Difference between Large and Small sample

Sr. No.	Large sample	Small sample
1.	The sample size is greater than 30.	The sample size is 30 or less than 30
2.	The value of a statistic obtain from the sample can be taken as an estimate of the population parameter.	The value of a statistic obtain from the sample can not be taken as an estimate of the population parameter.
3.	Normal distribution is used for testing.	Sampling distribution like t, F etc. are used for testing.

Test of Hypothesis (CO2)

❖ **Test of hypothesis for large samples(n>30)**

The following are the some important applications of test of hypothesis in case of large samples-

- Test of hypothesis about population mean
- Test of hypothesis about difference between two population mean

Test of hypothesis about population mean

let $\bar{x}$ be the mean of a large random sample of size n drawn from a normal population with mean μ and S.D. σ. To test the hypothesis that population mean μ has a specified value the test statistics to be used is given by

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Test of Hypothesis (CO2)

For large sample $\sigma \cong s$

Critical Value	Level of significance	
	1%	5%
Two tailed test	$\lvert Z \rvert = 2.58$	$\lvert Z \rvert = 1.96$
Single tailed test	$\lvert Z \rvert = 2.33$	$\lvert Z \rvert = 1.64$

Test of Hypothesis (CO2)

Example-1: A random sample of 900 members has a mean 3.4 cm. Can it be reasonably regarded as a sample from a large population of mean 3.2 cm and S.D. 2.3 cm?(Test at 5% level of significance)

Sol: Given $n = 900, \mu = 3.2, \sigma = 2.3, \bar{x} = 3.4$

H_0: $\mu = 3.2$

H_1: $\mu \neq 3.2$ (Two tailed test)

$\alpha = 0.05$

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ =0.261

Table value $Z = 1.96$

Conclusion : $|Z|_{cal} < table\ value$

So H_0 is accepted.

i.e sample comes out from population

Test of Hypothesis (CO2)

Example-2: An insurance agent has claimed that the avg. age of policy holders who insure through him is less then the avg. age of all agents which is 30.5 yrs.

A random sample of 100 policy holders who had insured through him gave the following age distribution-

Age	16-20	21-25	26-30	31-35	36-40
No. of person	12	22	20	30	16

Calculate arithmetic mean and standard deviation(S.D.) of this distribution and test his claim at 5% level of significance.

Test of Hypothesis (CO2)

Sol:

Age	No. of person(f)	x	$f.x$	$f(x-\bar{x})^2$
16-20	12	18	216	1399.68
21-25	22	23	506	740.08
26-30	20	28	560	12.8
31-35	30	33	990	529.2
36-40	16	38	608	1354.24
Total	100		2880	4036

$$\bar{x} = \frac{\sum f_i x_i}{N} = \frac{2880}{100} = 28.8; \quad s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}} = 6.352$$

Test of Hypothesis (CO2)

$H_0: \mu = 30.5$

$H_1:\ \mu < 30.5$ (Single tailed test)

$\alpha = 0.05$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = =-2.681$$

Table value:$Z = 1.64$

Conclusion : $|Z|_{cal} > table\ value$

So H_0 is rejected.

His claim is right.

Test of Hypothesis (CO2)

❖ **Test of hypothesis about difference between two population mean**

Let $\overline{x_1}$ be the mean of a sample of size $n_1$from a population with mean $\mu_1$and variance ${\sigma_1}^2$. Let $\overline{x_2}$ be the mean of a sample of size n_2 from a population with mean $\mu_2$and variance ${\sigma_2}^2$. To test whether the two population means are equal or not, the test statistic to be used as-

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

Test of Hypothesis (CO2)

Note :

- When $\sigma_1 = \sigma_2 = \sigma$ then $Z = \dfrac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

If σ is unknown and $\sigma_1 = \sigma_2$ then

$$\sigma^2 = \frac{n_1 {s_1}^2 + n_2 {s_2}^2}{n_1 + n_2}$$

- If σ_1, σ_2 are unknown & $\sigma_1 \neq \sigma_2$ then

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

Test of Hypothesis (CO2)

Example: The average income of persons was Rs 210 with S.D. of Rs 10 in a sample of 100 people of a city. For another sample of 150 persons the avg. income was Rs 220 with S.D. of Rs 12. Test whether there is any significance difference between the avg. income of the localities.(Test at 5% level of significance)

Sol:

Given $n_1 = 100, n_2 = 150, \overline{x_1} = 210,$
$\overline{x_2} = 220, s_1 = 10, s_2 = 12$
$H_0 \colon \mu_1 = \mu_2$
$H_1 \colon \mu_1 \neq \mu_2$

Test of Hypothesis (CO2)

$\alpha = 0.05$

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = -7.14$$

Table value Z = 1.96

Conclusion : $|Z|_{cal} > table\ value$

So H_0 is rejected.

Yes there is a significance difference the avg. income of localities.

P-value (CO2)

P-Value (Observed Significance Level):

It's the measure of the inconsistency between the hypothesized value for a population parameter and the observed sample statistic.

Using the P-Value

- Reject H_0 if $p - value \leq \alpha$
- Fail to Reject H_0 if $p - value > \alpha$

For a One-Tailed Test:

- Reject H_0 if the p -value is less than the significance level α.
- Accept H_0 otherwise.

For a Two-Tailed Test:

- Reject H_0 if the p -value is less than ½ of the significance level α. i.e. .025
- Accept H_0 otherwise.

P-value (CO2)

Example-1 A manufacturing process produces TV. tubes with an average life $\mu = 1200$ hr and σ $= 300\ hrs$. A new process is thought to give tubes a higher average life. And out of a sample of 100 tubes we find that they have an average life $\bar{x} = 1265$ hr. Is the new process really any better than the old?

Solution:

$H_0: \mu = 1200$

$H_1:\ \mu > 1200$(Single tailed test)

$\alpha = 0.05$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = 2.17$$

P-value (CO2)

$P(\bar{x} \geq 1265) = P(z \geq 2.17) = .015$

Here we reject H_0

Since $p - value \leq \alpha = 0.05$

Confidence Interval

We would like to determine an interval in which the population parameter is supposed to lie.

95% two sided C.L for μ is $\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \qquad \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right]$

99% two sided C.L for μ is $\left[\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \qquad \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right]$

Example: A random sample of 200 measurements from a large population gave a mean value of 50 and S.D. of 9. Determine 95% confidence interval for mean of population.

Sol: CI for μ:

Test of Hypothesis (CO2)

$$\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \qquad \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

$$[48.75 \qquad 51.24]$$

Test of Hypothesis (CO2)

Test of hypothesis for small samples ($n \leq 30$)

- t-test
- F-test

Test of hypothesis for Small Samples

When the size of the sample $n \leq 30$, then the sample is called small sample. For such sample it will not be possible for us to assume that the random sampling distribution of a statistic is approximately normal. In this sampling distribution is t-distribution.

Daily Quiz(CO2)

1. Write a short note on Null hypothesis?
2. Write a short note on Alternative hypothesis?
3. A sample of 800 students from a university was taken and their average weight was found to be 112 pounds with S.D. of 20 pounds. Could the mean weight in the population be 120 pounds.(Take $\alpha = 0.05$)

Recap(CO2)

- ✓ Test of hypothesis
- ✓ Procedure of Test of hypothesis
- ✓ Null Hypothesis
- ✓ Alternate Hypothesis
- ✓ Level of significance
- ✓ Test of hypothesis about difference between two population mean
- ✓ Test of hypothesis about Small samples

Topic objective (CO2)

t-Test

- Explanation why a t distribution is associated with $n - 1$ degrees of freedom and describe the information that is conveyed by the t statistic.
- Calculation the degrees of freedom for a one-sample t test and a two-independent-sample t test, and locate critical values in the t table.
- Identify the assumptions for the one-sample t test.
- Identify the assumptions for the two-independent-sample t test.

t-Test (CO2)

t-distribution: It is used when sample size is ≤ 30 and the population S.D. is unknown.

Application of t-distribution:

- To test the significance (hypothesis) of the difference between the mean of small sample and mean of the population
- To test the significance of the difference between the mean of two small samples.

t-Test (CO2)

❖ **To test the significance of the difference between the mean of small sample and mean of the population:**

When population is distributed normally and σ is unknown then statistic

$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ follows t-distribution with $n-1$ degree of freedom.

where $S = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$ = Calculated Standard deviation of sample

$\bar{x}$ →mean of sample

μ →mean of Population

n →sample size

t-Test (CO2)

Note-If standard deviation s of sample is given then

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

follows t−distribution with $n - 1$ degree of freedom.

$$\because \text{S} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \text{and } s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

From here we get

$$S = \sqrt{\frac{n}{n-1}}\, s$$

Example-1: The 9 items of a sample have the following values-
45,47,50,52,48,47,49,53,51

t-Test (CO2)

Does the mean of these values differ significantly from the assumed mean 47.5.(Test at 5% level of significance)

Sol: Here sample is small so we use t-test

x	45	47	50	52	48	47	49	53	51	$\sum x$ =442
$(x-\bar{x})^2$	16.81	4.41	0.81	8.41	1.21	4.41	0.01	15.21	3.61	$\sum(x-\bar{x})^2 = 54.89$

Now $S = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = 2.619$

$$\bar{x} = \frac{\sum \boldsymbol{x}}{n} = 49.1$$

t-Test (CO2)

H_0: $\mu = 47.5$

H_1: $\mu \neq 47.5$ (Two tailed test)

$\alpha = 0.05$

t= $\frac{\bar{x}-\mu}{s/\sqrt{n}} = 1.8327$

Degree of freedom=9-1=8

Table value: $t_{(8)}(0.05) = 2.306$

Conclusion : $|t|_{cal} < table\ value$

So H_0 is accepted.

There is no significant difference between their means.

t-Test (CO2)

Example-2: A sample of 20 items has mean 42 units and S.D. 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.

(Test at 5% level of significance)

Sol: here $n = 20, \bar{x} = 42, s = 5$

$H_0: \mu = 45$

$H_1: \mu \neq 45$ (Two tailed test)

$\alpha = 0.05$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = -2.61$$

Degree of freedom=20-1=19

Table value: $t_{(19)}(0.05) = 2.093$

t-Test (CO2)

Conclusion : $|t|_{cal} > table\ value$

So H_0 is rejected.

i.e. there is significance difference between sample mean & population mean.

Example-3: The height of 8 males participating in an athletic championship are found to be 175,168,165,170,167,160,173 and 168 cm. Can we conclude that the avg. height is greater than 165 cm.(Test at 5% level of significance)

Sol: Here S $= \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = 4.6521$

$\bar{x} = \frac{\sum x}{n}$ =168.25

t-Test (CO2)

$H_0: \mu = 165$

$H_1: \mu > 165$ (single tailed test)

$\alpha = 0.05$

t= $\frac{\bar{x}-\mu}{s/\sqrt{n}} = 1.97$

Degree of freedom=8-1=7

Table value: $t_{(7)}(0.05) = 1.895$

Conclusion : $|t|_{cal} > table\ value$

So H_0 is rejected.

i.e. avg. height is greater than 165.

Daily Quiz(CO2)

1. A sample of 1000 students from a university was taken and their average weight was found to be 112 pounds with S.D. of 20 pounds. Could the mean weight in the population be 120 pounds.(Take $\alpha = 0.05$)
2. Write a short note on Null hypothesis?
3. Samples of size 10 and 14 were taken from two normal populations with S.D. 3.5 and 5.2. The sample means were found to be 20.3 and 18.6. Test whether the means of two populations are the same. If the tabulated value is 2.0739 at 5% level.

t-Test (CO2)

To test the significance of the difference between the mean of two small samples.

Let two independent sample of size n_1 and n_2 be drawn from two normal population with mean $\mu_1$and μ_2 and equal S.D.$(\sigma_1 = \sigma_2 = \sigma)$but unknown. To test whether the two population mean are equal or whether the difference $\overline{x_1} - \overline{x_2}$ is significant we use t-test and the appropriate test statistic to be used is

$t = \dfrac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ with $n_1 + n_2 - 2$ degree of freedom.

t-Test (CO2)

Note:

- if the samples S.D. s_1 and s_2 are given then we have $S^2 = \frac{n_1 {s_1}^2 + n_2 {s_2}^2}{n_1 + n_2 - 2}$

So $S = \sqrt{\frac{n_1 {s_1}^2 + n_2 {s_2}^2}{n_1 + n_2 - 2}}$

- if s_1 and s_2 are not given

Then

$$S = \sqrt{\frac{\sum(x_1 - \overline{x_1})^2 + \sum(x_2 - \overline{x_2})^2}{n_1 + n_2 - 2}}$$

t-Test (CO2)

Example-1 Samples of sizes 10 and 14 were taken from two normal populations with S.D. 3.5 and 5.2. The sample means were found to be 20.3 and 18.6. Test that whether the means of two populations are the same at 5% level.

Sol:

Here $n_1 = 10, n_2 = 14, s_1 = 3.5, s_2 = 5.2$

$\overline{x_1} = 20.3, \overline{x_2} = 18.6$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$\alpha = 0.05$

t-Test (CO2)

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ where } S = \sqrt{\frac{n_1 {s_1}^2 + n_2 {s_2}^2}{n_1 + n_2 - 2}} = 4.772$$

$t = 0.8604$

Degree of freedom $= 10 + 14 - 2 = 22$

Table value: $t_{(22)}(0.05) = 2.07$

Conclusion : $|t|_{cal} < table\ value$

So H_0 is accepted.

Example-2: The height of 6 randomly sailors 63,65,68,69,71 and 72. Those of 9 randomly chosen soldiers are 61,62,65,66,69,70,71,72 and 73.

t-Test (CO2)

Test whether the sailors are on the avg. taller than soldiers. (Test at 5% level of significance)

Sol:

Let x_1 and x_2 be two samples denoting the height of sailors & soldiers.

Here $n_1 = 6, n_2 = 9$

Now $S = \sqrt{\frac{\sum(x_1-\overline{x_1})^2+\sum(x_2-\overline{x_2})^2}{n_1+n_2-2}} = 4.038,$ $\overline{x_1} = 68, \overline{x_2} = 67.66$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$ (Single tailed test)

$\alpha = 0.05$

t-Test (CO2)

x_1	x_2	$(x_1 - \overline{x_1})^2$	$(x_2 - \overline{x_2})^2$
63	61	25	44.3556
65	62	9	32.0356
68	65	0	7.0756
69	66	1	2.7556
71	69	9	1.7956
72	70	16	5.4756
	71		11.1556
	72		18.8356
	73		28.5156
Total = 408	609	60	152.0004

t-Test (CO2)

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ where } S = \sqrt{\frac{\sum(x_1 - \overline{x_1})^2 + \sum(x_2 - \overline{x_2})^2}{n_1 + n_2 - 2}} = 4.038$$

$t = 0.1597$

Degree of freedom $= 6 + 9 - 2 = 13$

Table value: $t_{(13)}(0.05) = 1.77$

Conclusion : $|t|_{cal} <$ table value

So H_0 is accepted.

i.e. the sailor are not on avg. taller than soldiers

t-Test (CO2)

Confidence limit for population mean

If t_α is the value of t at level of significance α at $n-1$ degree of freedom then

For two tailed test

H_0 is accepted if $t_{cal} <$ table value

$$\left|\frac{\bar{x}-\mu}{S/\sqrt{n}}\right| < t_\alpha$$

$$\Rightarrow \bar{x} - t_\alpha \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_\alpha \frac{S}{\sqrt{n}}$$

95% confidence limits(5% level of significance) are $\left[\bar{x} - t_{0.05}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}\frac{S}{\sqrt{n}}\right]$

99% confidence limits(1% level of significance) are $\left[\bar{x} - t_{0.01}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{0.01}\frac{S}{\sqrt{n}}\right]$

t-Test (CO2)

Example: A random sample of size 16 has 53 as mean. The sum of square of deviation from mean is 150. Find the two sided 95% confidence interval for population mean.

Sol:

$n = 16,\ \bar{x} = 53, \sum(x_i - \bar{x})^2 = 53$

$$S = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = 3.162$$

Degree of freedom(df)=16-1=15

$t_{15}(0.05) = 2.131$

95% confidence limits for μ are $\left[\bar{x} - t_{0.05}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}\frac{S}{\sqrt{n}}\right]$

$[51.315 \quad 54.685]$

Problem: Intelligent test were given to two groups of boys & girls

	Mean	S.D.	Size
Girl	75	8	60
Boys	73	10	100

Examine if the difference between mean scores is significant? (Take $\alpha = 0.05$)

Ans: H_0 is accepted

Weekly Assignment(CO2)

1 Intelligent test were given to two groups of boys & girls

	Mean	S.D.	Size
Girl	68	5	60
Boys	85	10	100

Examine if the difference between mean scores is significant? (Take $\alpha = 0.05$)

2. Define Null hypothesis for

i. t test

ii. Z-test

Recap(CO2)

- ✓ Test of hypothesis
- ✓ Procedure of Test of hypothesis
- ✓ Null Hypothesis
- ✓ Alternate Hypothesis
- ✓ Level of significance
- ✓ Test of hypothesis about difference between two population mean
- ✓ t-Test

Test of hypothesis for small samples ($n \leq 30$)

- t-test
- F-test

Test of hypothesis for Small Samples

When the size of the sample $n \leq 30$, then the sample is called small sample. For such sample it will not be possible for us to assume that the random sampling distribution of a statistic is approximately normal. In this sampling distribution is F-distribution.

F-Test (CO2)

- **F-Test:** In testing the Significance difference of two means of two Samples, we assumed that the two Samples came from the Same Population or population with equal Variance. The object of the F-test is to discover whether two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having the same variance.
- Hence before applying the t-test for the Significance of difference of two means, we have to test for the equality of population variance by using F-test.
- Let $n_1 \ and \ n_2$ be the Sizes of two Samples with Variances s_1^2 and s_2^2 respectively.

F-Test (CO2)

Case I : If s_1^2 & s_2^2 are given and $S_1^2 > S_2^2$.

Where $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ *and* $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$

Then Test Statistic $F = \frac{S_1^2}{S_2^2} > 1$ with degree of freedom is $\nu_1 = n_1 - 1, \nu_2 = n_2 - 1$.

And If $S_2^2 > S_1^2$ then

Test Statistic $F = \frac{S_2^2}{S_1^2} > 1$ with degree of freedom is

$\nu_2 = n_2 - 1, \nu_1 = n_1 - 1$

F-Test (CO2)

Case II : If s_1^2 & s_2^2 are not given.

Where $s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1}$ and $s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1}$

x_1 and x_2 are two Samples.

Hence Test Statistic $F = \frac{s_1^2}{s_2^2}$ $if\ s_1^2 > s_2^2$ with degree of freedom is $v_1 = n_1 - 1, v_2 = n_2 - 1$.

Or Test Statistic $= \frac{s_2^2}{s_1^2}$ $if\ s_2^2 > s_1^2$ with degree of freedom is $v_2 = n_2 - 1, v_1 = n_1 - 1$.

Conclusion: Same as t-test.

F-Test (CO2)

Q1. The two random samples reveal the following data:

Sample no.	Size	Mean	Variance
I	16	440	40
II	25	460	42

Test the equality of Population Variance.

Ans. For Checking the equality of Population Variance, we apply F-test.

Null Hypothesis $\boldsymbol{H_0}: \sigma_1^2 = \sigma_2^2$ i.e. there is no significant difference between the population Variances.

Alternative Hypothesis $\boldsymbol{H_1}: \boldsymbol{\sigma_1^2 \neq \sigma_2^2}$

Test Statistic: Since s_1^2 & s_2^2 are given then

F-Test (CO2)

Hence $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{16X\ 40}{15} = 42.7$

And $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{25X\ 42}{24}$=43.75

Since $S_2^2 > S_1^2$, Now $F = \frac{S_2^2}{S_1^2} = \frac{43.75}{42.7} = 1.02$ with degree of freedom is $\nu_2 = n_2 - 1, \nu_1 = n_1 - 1$ i.e.

$\upsilon_2 = 25 - 1 = 24\ and\ \upsilon_1 = 16 - 1 = 15$.

Conclusion: Here the calculate Value of $F = 1.02 < F_{(24,15)} = 2.29$ at 5% level of Significance. Hence H_0 is accepted. Therefore population variances are equal.

F-Test (CO2)

Example-2: The height of 6 randomly sailors 63,65,68,69,71 and 72. Those of 9 randomly chosen soldiers are 61,62,65,66,69,70,71,72 and 73. Test the equality of Population Variance. (Test at 5% level of significance)

Sol:

Let x_1 and x_2 be two samples denoting the height of sailors & soldiers.

Here $n_1 = 6, n_2 = 9$

Where $s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1}$ and $s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1}$

x_1 and x_2 are two Samples.

F-Test (CO2)

$\overline{x_1} = 68, \overline{x_2} = 67.66$

Null Hypothesis $\boldsymbol{H_0}: \sigma_1^2 = \sigma_2^2$ i.e. there is no significant difference between the population Variances.

Alternative Hypothesis $\boldsymbol{H_1}: \boldsymbol{\sigma_1^2 \neq \sigma_2^2}$

$\boldsymbol{Level\ of\ significance}\ \alpha = 0.05$

F-Test (CO2)

x_1	x_2	$(x_1 - \overline{x_1})^2$	$(x_2 - \overline{x_2})^2$
63	61	25	44.3556
65	62	9	32.0356
68	65	0	7.0756
69	66	1	2.7556
71	69	9	1.7956
72	70	16	5.4756
	71		11.1556
	72		18.8356
	73		28.5156
Total = 408	609	60	152.0004

F-Test (CO2)

$$s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{60}{5} = 12 \text{ and}$$

$$s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{152.0004}{8} = 19.00005$$

Test Statistic F test $= \frac{s_2^2}{s_1^2} = \frac{19.00005}{12} = 1.583375$

$if\ s_2^2 > s_1^2$ with degree of freedom is

$v_2 = n_2 - 1 = 9 - 1 = 8, v_1 = n_1 - 1 = 6 - 1 = 5.$

Table value: $F_{(8,5)}(0.05) = 4.82$

Conclusion : $|F|_{cal} <$ table value

So H_0 is accepted.

i.e. Therefore population variances are equal.

Daily Quiz(CO2)

Problem: The random samples were drawn from two normal populations and the following results were obtained

Sample 1: 16,17,18,19,20,21,22,24,26,27

Sample 2: 19,22,23,25,26,28,29,30,31,32,35,36

Obtain estimates of the variances of populations and test whether the two populations have the same variances.

(Take $\alpha = 0.05$)

Topic Objective (CO2)

Chi Square Test(χ^2)

- Distinguish between nonparametric and parametric tests.
- Explanation how the test statistic is computed for a chi-square test.
- Compute the chi-square goodness-of-fit test and interpret the results.
- Identify the assumption and the restriction of expected frequency size for the chi-square test.
- Compute the chi-square test for independence and interpret the results.

Chi Square Test(χ^2)(CO2)

Chi Square Test(χ^2)

1. Chi-square test of goodness of Fit.
2. Chi-square test of independence of Attributes.

Chi Square Test(χ^2)(CO2)

Chi Square Test(χ^2)

- Chi Square test is the test of significance.
- Chi square test is a useful measures of comparing experimentally obtained result with those expected theoretically and based on hypothesis.
- $$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$
- It is a mathematical expression, representing the ratio of experimentally obtained result(O) and theoretically expected result(E) based on certain hypothesis. It uses data in form of frequencies.

Chi Square Test(χ^2)(CO2)

- Chi square test is calculated by dividing the square of overall deviation in the observed and expected frequencies by expected frequency. $\boldsymbol{\chi^2 = \sum \frac{(O-E)^2}{E}}$

O=the frequencies observed, E=the frequencies expected, Σ= sum of

- If there is no difference between actual and observed frequencies, the value of Chi- Square is zero.
- If there is difference between actual and observed frequencies, the value of Chi- Square would be more than zero.
- But the difference in the observed frequencies, may also be due to the sampling fluctuations and it should be ignored in drawing interference.

Degree of Freedom

- In test, while comparing the calculated value with the table value, we have to calculate the degree of freedom. The degree of freedom is calculated from the no of classes. Therefore, the no. of degree of freedom in a test is equal to the no. of classes minus one.
- If there are two classes, three classes and four classes, the degree of freedom would be 2-1,3-1 and 4-1, respectively. In a contingency table, the degree of freedom is calculated in different manner:

$$d.f.=(r-1)(c-1)$$

Where r= no. of row in the table,C= no. of column in the table

- Thus in 2×2 contingency table, the degree of freedom is (2-1)(2-1)=1. Similarly, in a 3×3 contingency table, the degree of freedom is

Chi Square Test(χ^2)(CO2)

(3-1)(3-1)=4. :Likewise in 3×4 contingency table degree of freedom is (3-1)(4-1)=6, and so on.

Example: 2 Rows × 2 columns

	COLUMN 1	COLUMN 2	ROW TOTAL
ROW 1	+	+	RT 1
ROW 2	+	+	RT 2
COLUMN TOTAL	CT 1	CT 2	

Degree of freedom = (r-1)(c-1)=(2-1)(2-1)=1

Chi Square Test(χ^2)(CO2)

Application of chi square test

- The chi square are applicable in varied problems in agriculture, biology engineering and medical sciences.
- To test the goodness of fit.
- To test the independence of attributes.
- To test the detection of linkage.
- To test the homogeneity of independent estimates of the population variance.

Chi Square Test(χ^2)(CO2)

Example: The demand for a particular spare part in a factory was found to vary from day-to-day. In a sample study the following information was obtained:

Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
No. of parts demanded	1124	1125	1110	1120	1126	1115

Under the null hypothesis, the expected frequencies of the spare part demanded on each of the six days would be:(1124+1125+1110+1120+1126+1115)/6=6270/6=1120

Chi Square Test(χ^2)(CO2)

Days	Observed (O_i)	Expected (E_i)	(O_i-E_i)	$(O_i - E_i)^2$
Monday	1124	1120	4	16
Tuesday	1125	1120	5	25
Wednesday	1110	1120	-10	100
Thursday	1120	1120	0	0
Friday	1126	1120	6	36
Saturday	1115	1120	-5	25
total	6720	6720		

Chi Square Test(χ^2)(CO2)

$$\chi^2 = \sum_{i=1}^{n} \frac{(Observed - Expected)^2}{Expected} = 0.179$$

The number of degrees of freedom = 6 –1 = 5 (since we are given 6 frequencies subjected to only one linear constraint: $\sum O_i = \sum E_i = 6720$).
The tabulated $\chi^2{}_{0.05}$ for 5 d.f.= 11.07

Conclusion: Since calculated value of χ^2 is less than the tabulated value, it is not significant and the null hypothesis may accepted at 5% level of significance. Hence we conclude that the numbers of parts demanded are same over the 6-day period.

Chi Square Test(χ^2)(CO2)

Chi square test of Independent of Attributes:

- Let us consider two attributes A and B divides into r classes $A_{1,}A_2, A_3 \ldots A_r$and B divided into s classes $B_{1,}B_2, B_3 \ldots B_s$.
- Such a classification in which attributes are divided into more than two classes is known as manifold classification.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{[(A_iB_j) - (A_iB_j)_o]^2}{(A_iB_j)_o} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{[O_{ij} - E_{ij}]^2}{E_{ij}}$$

Chi Square Test(χ^2)(CO2)

Where $\boldsymbol{O_{ij}}$=$\boldsymbol{A_i B_j}$: observed frequency for contingency table category in column i and row j.

$\boldsymbol{E_{ij}}$=$\boldsymbol{(A_i B_j)_o}$: expected frequency for contingency table category in column i and row j.

$\boldsymbol{(A_i B_j)_o}=\frac{(A)_i (B)_j}{N}$, (i=1,2,3…r,j=1,2,3…s)

Which is distributed as a $\boldsymbol{\chi^2}$with (r-1)(s-1) d.f.

Example : Two sample polls of votes for two candidates A and B for public office are taken, one from among the resident area of ruler areas. The results are given in the following table. Examine whether the nature of area related to the voting preference in elections.

Chi Square Test(χ^2)(CO2)

Area	Voter for A	Voter for B	Total
Ruler	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

Solution : H_0: the nature of the area is independent of the voting preference in the election

Expected frequencies

E(620)=$\frac{1170\times1000}{2000} = 585$

E(380)=$\frac{830\times1000}{2000} = 415$

E(550)=$\frac{1170\times1000}{2000} = 585$

Chi Square Test(χ^2)(CO2)

E(450)=$\frac{830\times1000}{2000} = 415$

Calculate the statistic $\boldsymbol{\chi^2}$

$=\frac{(620-585)^2}{585}+\frac{(380-415)^2}{415}+\frac{(550-585)^2}{585}+\frac{(450-415)^2}{415}$

=10.0891

The tabulated $\boldsymbol{\chi^2}_{0.05}$ for (2-1)(2-1) d.f.= 3.841

Conclusion: since calculate $\boldsymbol{\chi^2}$ is much greater than the tabulated value, it is highly significant and null hypothesis is rejected at 5% of level of significance. Thue we conclude that nature of area is rejected to voting preference in the election.

Chi Square Test(χ^2)(CO2)

Example : What are the expected frequencies of the contingency tables given below:

Observed frequency:

p	q	p+q
r	s	r+s
p+r	q+s	p+q+r+s

Expected frequency:

$\frac{(p+r) \times (p+q)}{(p+q+r+s)}$	$\frac{(q+s) \times (p+q)}{(p+q+r+s)}$
$\frac{(p+r) \times (r+s)}{(p+q+r+s)}$	$\frac{(q+s) \times (r+s)}{(p+q+r+s)}$

Chi Square Test(χ^2)(CO2)

Example : To test the effectiveness of inoculation against cholera, the following table was obtained:

	Attacked	Not attacked	Total
inoculated	30	160	190
Not inoculated	140	460	600
Total	170	620	790

(figures represent the number of persons.)

Use χ^2 test to defend or refute the statement that the inoculation prevents attack from cholera.

Solution : H_0:the inoculation does not prevents attack from cholera.

Chi Square Test(χ^2)(CO2)

Expected frequencies

	Attacked	Not attacked
inoculated	$\frac{190 \times 170}{790} = 40.886$	$\frac{190\times620}{790}$ =149.11
Not inoculated	$\frac{600\times170}{790}$ =129.11	$\frac{600\times620}{790}$ =470.89

Calculate the statistic χ^2

Oi	30	160	140	690
Ei	40.886	149.11	129.11	470.89
$\frac{(Oi - Ei)^2}{Ei}$	2.898	0.795	0.918	0.252

Chi Square Test(χ^2)(CO2)

$\boldsymbol{\chi}^{\mathbf{2}}{}_{cal} = \sum \frac{(\boldsymbol{Oi-Ei})^{\mathbf{2}}}{\boldsymbol{Ei}}$ =4.8863

The tabulated $\boldsymbol{\chi}^{\mathbf{2}}{}_{0.05}$for (2-1)(2-1) d.f.= 3.841

Conclusion: since calculate $\boldsymbol{\chi}^{\mathbf{2}}$is much greater than the tabulated value, it is highly significant and null hypothesis is rejected at 5% of level of significance. Hence we defend the statement that inoculation prevents attack from cholera.

Daily Quiz(CO2)

Q1: A sample analysis of examination results of 500 students it was found that 220 students have failed, 170 have secured a third class, 90 have secured a second class and the rest, a first class. Do these figures support the general belief that above categories are in the ratio 4:3:2:1 respectively?(Test at 5 % level of significance)

Q2: A sample of 400 students of under graduate and 400 students of post graduate classes was taken to know their opinion about autonomous colleges.290 of the under graduate and 310 of the post graduate students favored the autonomous status. Present these facts in the form of a table and test at 5% level that the opinion regarding autonomous status of colleges is independent of the level of classes of students.

Recap(CO2)

- ✓ Test of hypothesis
- ✓ Test of hypothesis about difference between two population mean
- ✓ To test the significance of the difference between the mean of two small samples.
- ✓ χ^2 test

Topic objective (CO2)

Analysis of variance(ANOVA)

- Appropriately interpret results of **analysis of variance** tests.
- Distinguish between one and two factor **analysis of variance** tests.
- Identify the appropriate hypothesis testing procedure based on type of outcome variable and number of samples.

Analysis of Variance (ANOVA)(CO2)

ANOVA:

- ANOVA is used to determine there is any significant difference between the mean of three or more independent samples.
- ANOVA is carried out on the basis of ratio between two variances, this ratio forms the test statistic known as F-statistics given by-

$$F = \frac{\text{variance between the samples}}{\text{variance with in the sample}}$$

Analysis of Variance (ANOVA)(CO2)

ANOVA is to split up the total variation as follows:

1. Variance between samples.
2. Variance within samples.

ANOVA classification

1. One way ANOVA: This is used to see the effect of an independent variable on a given dependent variable.
2. Two way ANOVA: This is used to see the effect of two independent variable on a given dependent variable.

Analysis of Variance (ANOVA)(CO2)

ANOVA TABLE:

SOURCE OF VARIATION	SUM OF SQUARES	DEGREE OF FREEDOM	MEAN SUM OF SQUARES	VARIANCE RATIO OR F
BETWEEN SAMPLES (COLUMN)	(SSC)	c-1	MSC=$\frac{SSC}{c-1}$	F =$\frac{MSC}{MSE}$
WITH IN SAMPLES (ERROR)	(SSE)	n-c	MSE=$\frac{SSE}{n-c}$	
TOTAL	SST	n-1		

Computation of test statistic

$$F = \frac{\text{variance between the samples}}{\text{variance within the samples}} = \frac{MSC}{MSE}$$

Analysis of Variance (ANOVA)(CO2)

Q.1 It Is desired to compare three Hospitals with regards to number of deaths per month. A sample of death records of each hospital and the number of deaths was given below. From these data suggest a difference in the no. of the deaths per month among three hospitals.

A	B	C
3	6	7
4	3	3
3	3	4
5	4	6
0	4	5

Analysis of Variance (ANOVA)(CO2)

Sol.Null hypothesis H_0 :There is no difference in the no. of deaths per months among three hospitals.

Alternate hypothesis H_1:There is a significant difference in the no. of death per month among three hospitals.

Level of Significance : We use 5% level of Significance.

Test Statistic : To find the Variance ratio F, we set up an ANOVA tables as follows:

Sample totals:

$$\sum y_a = 3 + 4 + 3 + 5 + 0 = 15$$

$$\sum y_b = 6 + 3 + 3 + 4 + 4 = 20$$

$$\sum y_c = 7 + 3 + 4 + 6 + 5 = 25$$

Analysis of Variance (ANOVA)(CO2)

1. Grand total(T)=$\sum y_a + \sum y_b + \sum y_c$=60
2. Correction factor(C.F)=$\frac{(T)^2}{n} = \frac{60^2}{15} = 240$

Sum of Squares of Samples:

$$\sum {y^2}_A = 3^2 + 4^2 + 3^2 + 5^2 + 0^2 = 59$$

$$\sum {y^2}_B = 6^2 + 3^2 + 3^2 + 4^2 + 4^2 = 86$$

$$\sum {y^2}_C = 7^2 + 3^2 + 4^2 + 6^2 + 5^2 = 135$$

Total Sum of Squares= $\sum {y^2}_A + \sum {y^2}_B + \sum {y^2}_C$ - C.F

=59+86+135-240=40

Analysis of Variance (ANOVA)(CO2)

Sum of Squares between Samples

$$=\frac{(\sum y_A)^2}{n_1}+\frac{(\sum y_B)^2}{n_2}+\frac{(\sum y_C)^2}{n_3}-\frac{(T)^2}{n}$$

$$=\frac{(15)^2}{5}+\frac{(20)^2}{5}+\frac{(25)^2}{5}-240$$

= 10

Sum of squares with in the samples

= Total sum of squares – sum of squares between the samples

= 40-10

=30

Analysis of Variance (ANOVA)(CO2)

Degree of freedom for total sum of squares =n-1=15-1=14

Degree of freedom for Hospitals=c-1 = 3-1=2

Degree of freedom for Error =n-c =15-3=12

ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREE OF FREEDOM	MEAN SUM OF SQUARES	VARIANCE RATIO OR F
BETWEEN SAMPLES (COLUMN)	10(SSC)	2	$MSC=\frac{SSC}{c-1}=\frac{10}{2}=5$	$F=\frac{MSC}{MSE}=\frac{5}{2.5}=2$
WITH IN SAMPLES (ERROR)	30(SSE)	12	$MSE=\frac{SSE}{n-c}=\frac{30}{12}$ =2.5	
TOTAL	40	14		

Analysis of Variance (ANOVA)(CO2)

The tabular value of F at 5 % level of significance with degree of freedom ν_1=2 & $\nu_2 = 12\ is\ 3.89$:

Conclusion: $F_{cal} < F_{tab}$, i. e. H_0 is accepted, so the difference is insignificant and we conclude that the data do not suggest a difference in the no of deaths per month among the three hospitals.

Analysis of Variance (ANOVA)(CO2)

Q.2 A manufacturing company purchased three new machines of different makes and wishes to determine whether one of them is faster than the others in producing a certain out put . Five hourly production figures are observed at random from each machine and the results are given below:

Observations	A1	A2	A3
1	25	31	24
2	30	39	30
3	36	38	28
4	38	42	25
5	31	35	28

Analysis of Variance (ANOVA)(CO2)

Use ANOVA and determine whether the machines are significantly Different in their mean speed.(Given at 5% level,F2,12=3.89)

Sol. Null hypothesis H0: Machines are not significantly different in their mean speed i.e.,$\mu 1 = \mu 2 = \mu 3$.

Alternate hypothesis H_1:Machines are significantly different in their mean speed.

Level of Significance: We use 5% level of Significance.

Test Statistic: T o find the Variance ratio F, we set up an ANOVA tables as follows: let us shift the origin at 30 i.e., reduce each observations by 30.we have

Analysis of Variance (ANOVA)(CO2)

Sample totals:

$$\sum A_1 = -5 + 0 + 6 + 8 + 1 = 10$$

$$\sum A_2 = 1 + 9 + 8 + 12 + 5 = 35$$

$$\sum A_3 = -6 + 0 - 2 - 5 - 2 = -15$$

Observations	A1	A2	A3
1	-5	1	-6
2	0	9	0
3	6	8	-2
4	8	12	-5
5	1	5	-2

Analysis of Variance (ANOVA)(CO2)

1. Grand total(T)=$\sum A_1+\sum A_2+\sum A_3$=30

2. Correction factor(C.F)=$\frac{(T)^2}{n}=\frac{(30)^2}{15}=60$

Sum of Squares of Samples:

$$\sum {A^2}_1=(-5)^2+0+6^2+8^2+1^2=126$$

$$\sum {A^2}_2=1^2+9^2+8^2+12^2+5^2=315$$

$$\sum {A^2}_3=(-6)^2+0^2+(-2)^2+(-5)^2+(-2)^2=69$$

3. Total Sum of Squares= $\sum {A^2}_1+\sum {A^2}_2+\sum {A^2}_3$ - C.F

=126+315+69-60=450

Analysis of Variance (ANOVA)(CO2)

Sum of Squares between Samples

$$=\frac{(\sum A_1)^2}{n_1}+\frac{(\sum A_2)^2}{n_2}+\frac{(\sum A_3)^2}{n_3}-\frac{(T)^2}{n}$$

$$=\frac{(10)^2}{5}+\frac{(35)^2}{5}+\frac{(-15)^2}{5}-60$$

= 250

Sum of squares with in the samples

= Total sum of squares – sum of squares between the samples

= 450-250

=200

Degree of freedom for total sum of squares =n-1=15-1=14

Degree of freedom for Machines=c-1 = 3-1=2

Analysis of Variance (ANOVA)(CO2)

Degree of freedom for Error =n-c =15-3=12

ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREE OF FREEDOM	MEAN SUM OF SQUARES	VARIANCE RATIO OR F
BETWEEN SAMPLES (COLUMN)	250(SSC)	2	$MSC=\frac{SSC}{c-1}=\frac{250}{2}$ =125	$F_{2,12}=\frac{MSC}{MSE}=\frac{125}{16.67}=7.498$
WITH IN SAMPLES (ERROR)	200(SSE)	12	$MSE=\frac{SSE}{n-c}=\frac{200}{12}$ =16.67	
TOTAL	450	14		

Analysis of Variance (ANOVA)(CO2)

The tabular value of F at 5 % level of significance with degree of freedom ν_1=2 & $\nu_2 = 12$ *is* 3.89(given)

Conclusion: $F_{cal} > F_{tab},$ so the null hypothesis is rejected and the difference is significant and we conclude that there is significant difference in the mean speed of the machines.

Two way ANOVA(CO2)

- In two way classification, the data are classified according to two factor for example- the production of a manufacturing process can be studied on the basis of workers as well as machines. In two way classification the following procedure is adopted in the analysis of variance-

B / A	B_1	B_2	……	B_h	Total
A_1	X_{11}	X_{12}	……	X_{1h}	R_1
A_2	X_{21}	X_{22}	……	X_{2h}	R_2
⋮	⋮	⋮	⋮	⋮	⋮
A_k	X_{k1}	X_{k2}	……	X_{kh}	R_k
Total	C_1	C_2	……	C_h	

Two way ANOVA(CO2)

- $H_0: \mu_{A_1} = \mu_{A_2} = \cdots\cdots = \mu_{A_k}$ and $\mu_{B_1} = \mu_{B_2} = \cdots\cdots = \mu_{B_k}$
  $H_1: \mu_{A_1} \neq \mu_{A_2} \neq \cdots\cdots = \mu_{A_k}$ and $\mu_{B_1} \neq \mu_{B_2} \neq \cdots\cdots \neq \mu_{B_k}$
- Compute grand total of all observations i.e.
  $G = \sum_i \sum_j X_{ij}$
- Correction factor $(\mathrm{C.F.}) = \frac{G^2}{n}$ where n is total no. of items in all observations.
- Compute raw sum of square $\mathrm{RSS} = \sum_i \sum_j X_{ij}^2$
- Total sum of square, $\mathrm{SST} = \mathrm{RSS} - \mathrm{C.F.}$
- Sum of square between rows,
  $$\mathrm{SSR} = \frac{R_1^2 + R_2^2 + \cdots\cdots + R_k^2}{h} - C.F.$$
  here h is no. of item in each row.

Two way ANOVA(CO2)

- Sum of square between column,

  $$\text{SSC} = \frac{C_1^2 + C_2^2 + \cdots\cdots + C_h^2}{k} - C.F.$$

  here k is no. of item in each column.
- Sum of square of the error, $\text{SSE} = \text{SST} - \text{SSR} - \text{SSC}$

Two way ANOVA(CO2)

- **ANOVA Table:**

Source of variation	Sum of squares	Degree of freedom	Mean sum of squares	*F*-ratio
Between rows	SSR	$k-1$	$\text{MSSR} = \dfrac{\text{SSR}}{k-1}$	$F_R = \dfrac{\text{MSSR}}{\text{MSSE}}$
Between column	SSC	$h-1$	$\text{MSSC} = \dfrac{\text{SSW}}{h-1}$	$F_C = \dfrac{\text{MSSC}}{\text{MSSE}}$
Error	SSE	$(k-1)(h-1)$	$\text{MSSE} = \dfrac{\text{SSE}}{(k-1)(h-1)}$	

- If $F_{\text{cal}} <$ table value we accept Null Hypothesis H_0.
- If $F_{\text{cal}} >$ table value we reject Null Hypothesis H_0.

Example-1: The following table represent the no. of units of a commodity produced by 3 different workers using 3 different machines-

Workers / Machines	*A*	*B*	*C*
X	16	64	40
Y	56	72	56
Z	12	56	28

Test-(i) whether the mean productivity is the same for the different machines types.

(ii) Whether the three workers differ with regard to mean productivity.

Two way ANOVA(CO2)

- $H_0: \mu_A = \mu_B = \mu_C \ \& \ \mu_X = \mu_Y = \mu_Z$
  $H_1: \mu_A \neq \mu_B \neq \mu_C \ \& \ \mu_X \neq \mu_Y \neq \mu_Z$
- Compute grand total of all observations ,$G = \sum_i \sum_j X_{ij} = 400$
- Correction factor $(\mathrm{C.F.}) = \frac{G^2}{n} = 17777.78$
- Compute raw sum of square $\mathrm{RSS} = \sum_i \sum_j X_{ij}^2 = 21472$
- Total sum of square, $\mathrm{SST} = \mathrm{RSS} - \mathrm{C.F.} = 3694.22$
- Sum of square between rows,

  $$\mathrm{SSR} = \frac{R_1^2 + R_2^2 + \cdots\cdots + R_k^2}{h} - C.F. = 1379.55$$
- Sum of square between column,

  $$\mathrm{SSC} = \frac{C_1^2 + C_2^2 + \cdots\cdots + C_h^2}{k} - C.F. = 1987.55$$
- Sum of square of the error, $\mathrm{SSE} = \mathrm{SST} - \mathrm{SSR} - \mathrm{SSC} = 81.78$

Two way ANOVA(CO2)

- **ANOVA Table:**

Source of variation	Sum of squares	Degree of freedom	Mean sum of squares	F-ratio
Between rows	SSR $= 1379.55$	$k-1=2$	$\text{MSSR} = \frac{\text{SSR}}{k-1} = 993.7$	$F_R = \frac{\text{MSSR}}{\text{MSSE}} = 12.15$
Between column	SSC $= 1987.55$	$h-1=2$	$\text{MSSC} = \frac{\text{SSW}}{h-1} = 689.7$	
Error	SSE $= 327.12$	$(k-1)(h-1) = 4$	$\text{MSSE} = \frac{\text{SSE}}{(k-1)(h-1)} = 81.78$	$F_C = \frac{\text{MSSC}}{\text{MSSE}} = 8.43$

- Here $F_R > F_{(2,4)}(0.05) = 6.94$ & $F_C > F_{(2,4)}(0.05) = 6.94$

  We reject H_0.

Daily Quiz(CO2)

Q1: Three Variety A, B and C of wheat are shown in 5 plots each and the following yields per acre were obtained-

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

Set up a table of analysis of variance and find out whether there is any significant difference between the mean yields of these varieties.(Test at 5% level of significance)

Recap(CO2)

- ✓ Test of hypothesis
- ✓ Test of hypothesis about difference between two population mean
- ✓ To test the significance of the difference between the mean of two small samples.
- ✓ χ^2 test
- ✓ Anova

Topic objective (CO2)

Control Charts

- That process variable can be plotted on a **control chart** over time. The **objective** of the **control chart** is to find any "special" causes of variation as well as to reflect the process improvements that have been made.

Statistical Quality Control (SQC)

SQC refers to the use of statistical methods to monitoring & maintaining the quality of product & services.

Method of SQC:

SQC are applied to two distinct phase of plant operation, they are

1. Process control
2. Product control

1. Process control:

- Under the process control, the quality of the product is controlled while the product are in process of production. It is based on prob. Theoretically, It is common that when several identical parts are manufactured some are little large & some are little small but most will be of same size when the frequency or count of the items by size is plotted with size on the horizontal scale and count on the vertical scale a normal curve is obtained. The process control is secured with the technique of **control charts**.

2. Product control:

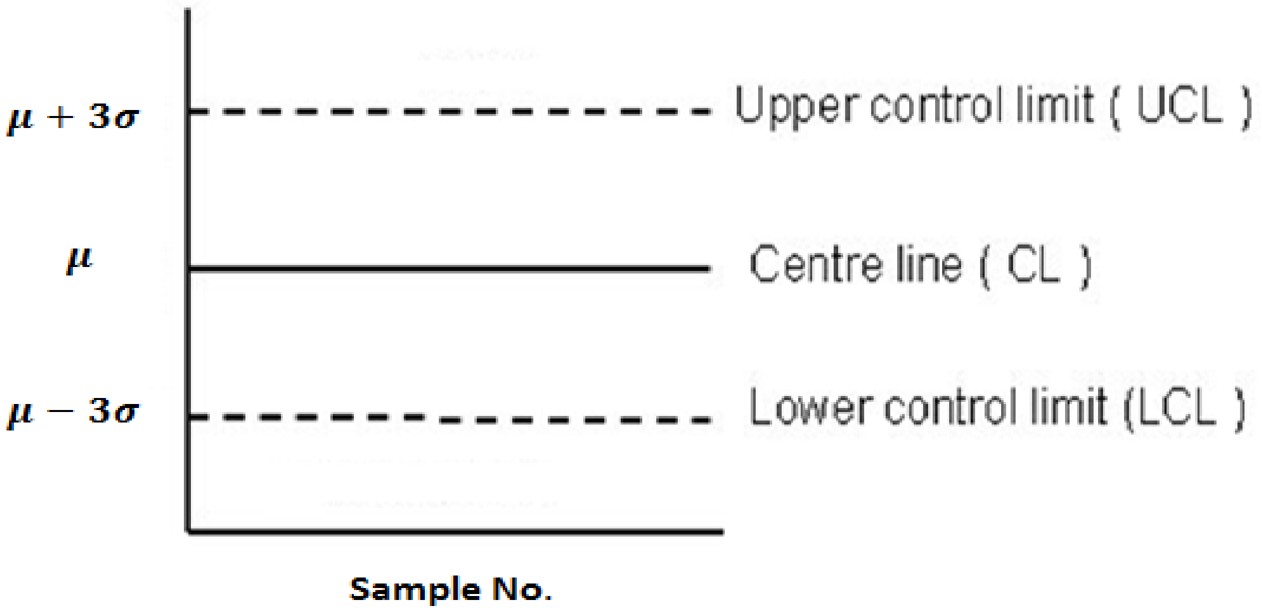
- Under the product control, the quality of the product is controlled while the product is ready for sale and dispatch to the customers.
- The product control is secured with the technique of acceptance sampling. Acceptance sampling is based on premises that a sample represent the whole lot from which the sample is drawn.
- In this method samples are taken out & are carefully inspected to detect defects. On the basis of defect found, the lot is accepted or rejected. If defect are few, lot is accepted. It is rejected when defect are more. Thus acceptance sampling is used to take decision regarding acceptance or rejection a lot without examine the entire lot.

Control Charts (CO2)

Control charts: Control chart is a graphical chart for detecting unnatural pattern of variation in the production process and determining the permissible limits of variation. There are 3 horizontal line in control chart-

1. **Control line(CL):** it passes through the middle of the chart and parallel is to the base. It represent the prescribed quality of product.
2. **Upper Control limit(UCL):** It is shown in the chart by a dotted line that passes through the chart above and parallel to the central line and represent upper limit of tolerance.
3. **Lower Control limit(LCL):** It is shown in the chart by a dotted line that passes through the chart below and parallel to the central line and represent lower limit of tolerance.

Control Charts (CO2)



Note:

- **If LCL is negative then we take LCL=0.**
- **Control chart are based on 3 sigma limits.**

Control Charts (CO2)

Control chart for variables- are used to monitor characteristic that can be measured for e.g- length, weight, diameter, time etc. such charts are of two types-

1. Mean chart ($\bar{x}$ chart)
2. Range chart (R chart)

Control Charts (CO2)

1. **Mean chart ($\overline{x}$ chart):** It is used to monitor the change in mean of a process.

Construction of $\overline{x}$ chart:

- Compute the mean of each sample.

  i.e. $\overline{x_1}, \overline{x_2} \ldots \ldots \overline{x_k}$ (here k is no. of sample)

- Compute the mean of sample mean.

  i.e. $\bar{\bar{x}} = \frac{\overline{x_1}+\overline{x_2}+\cdots\cdots+\overline{x_k}}{k}$

- Compute the range:

  $R =$ (largest value$-$smallest value) of each sample i.e. $R_1, R_2, \ldots \ldots R_k$.

Control Charts (CO2)

- Compute the mean of sample range

i.e. $\bar{R} = \frac{R_1+R_2+\cdots\cdots+R_k}{k}$

- CL= $\bar{\bar{x}}$

UCL= $\bar{\bar{x}} + A_2\bar{R}$

LCL= $\bar{\bar{x}} - A_2\bar{R}$

where A_2 are given in the table and depends upon the sample size.

- If the sample mean of each sample lie with in the UCL and LCL then process is in control. If it lie outside then process is out of control.

Control Charts (CO2)

2. Range chart(R-chart)- It is used to monitor the dispersion or variability of the process.

Construction of R-chart-

- Compute the range:

  $R =$ (largest value$-$smallest value) of each sample i.e. $R_1, R_2, \ldots \ldots R_k$.

- Compute the mean of sample range

  i.e. $\bar{R} = \frac{R_1+R_2+\cdots\cdots+R_k}{k}$

- CL$= \bar{R}$

  UCL$= D_4\bar{R}$

  LCL$= D_3\bar{R}$

Control Charts (CO2)

where D_4, D_3 are constant whose value depends upon the sample size and coming from table.

- If the range of each sample lie with in UCL and LCL then process is in control. If it lie outside then process is out of control.

Example-1: Construct mean chart ($\bar{x}$ chart) and range chart(R-chart) for the following data of 5 samples with each set of 5 items-(conversion factor for $n = 5, A_2 = 0.577,$

$D_3 = 0, D_4 = 2.115$)

Control Charts (CO2)

Sample No.	Weights				
1	20	15	10	11	14
2	12	18	10	8	22
5	21	19	17	10	13
4	15	12	19	14	20
5	20	19	26	12	23

Control Charts (CO2)

Sol:

Sample No.	Weights					Sample mean	Range R=L−S
1	20	15	10	11	14	14	10
2	12	18	10	8	22	14	14
5	21	19	17	10	13	16	11
4	15	12	19	14	20	16	8
5	20	19	26	12	23	20	14
Total						80	57

$\overline{x}$ chart:

$$\overline{x_1} = \frac{70}{5} = 14 \text{ similarly } \overline{x_2} = 14, \overline{x_3} = 16, \overline{x_4} = 16, \overline{x_5} = 20$$

Control Charts (CO2)

$R_1 = largest - smallest = 20 - 10 = 10$

$R_2 = 22 - 8 = 14$

$R_3 = 11$

$R_4 = 8$

$R_5 = 14$

$\bar{\bar{x}} = \frac{\overline{x_1}+\overline{x_2}+\cdots\cdots+\overline{x_5}}{5} = \frac{80}{5} = 16$

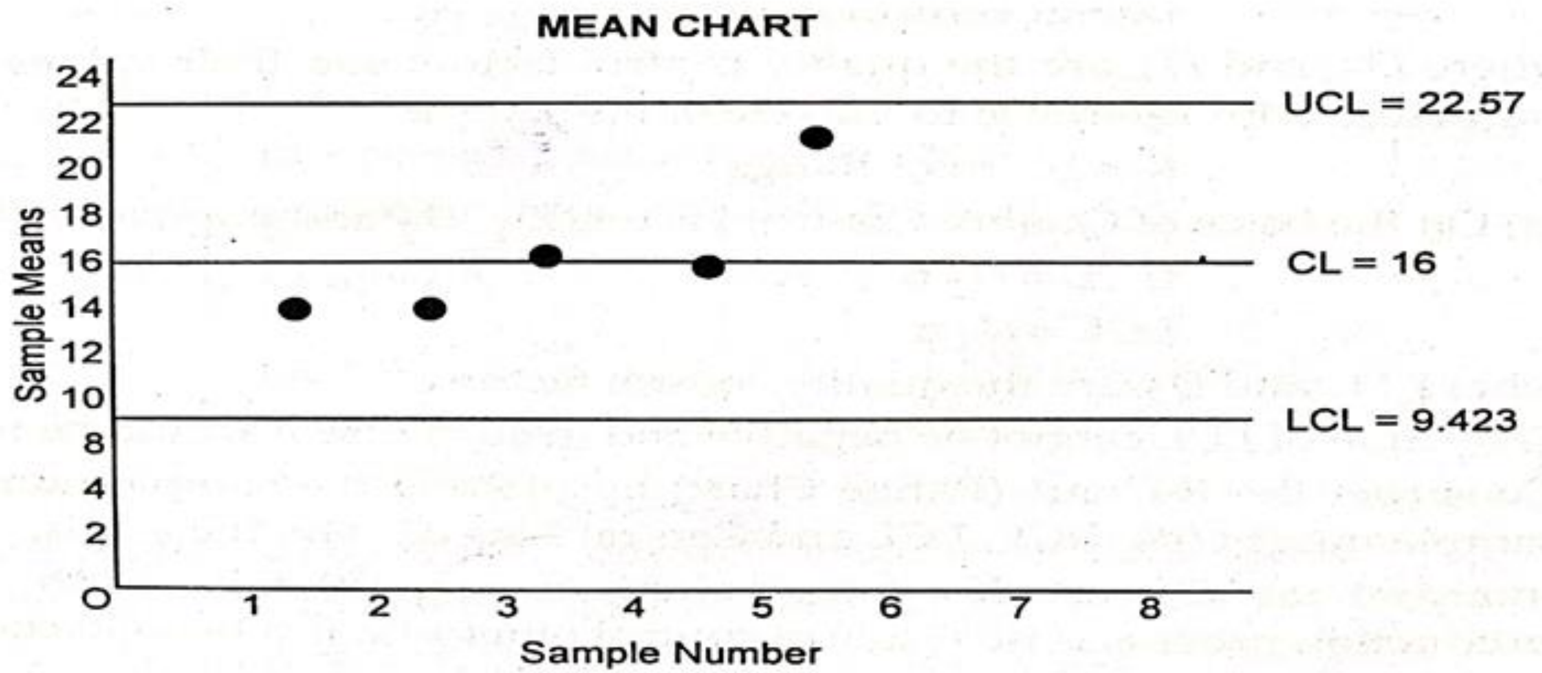
$\bar{R} = \frac{R_1+R_2+\cdots\cdots+R_5}{5} = \frac{57}{5} = 11.4$

- CL= $\bar{\bar{x}} = 16$

  UCL= $\bar{\bar{x}} + A_2\bar{R} = 22.577$ here $A_2 = 0.577$ (Given) LCL= $\bar{\bar{x}} - A_2\bar{R} = 9.423$

Control Charts (CO2)

Mean Chart-



As all the sample mean values fall within the control limits, the chart shows that the given process is in staistical control.

Control Charts (CO2)

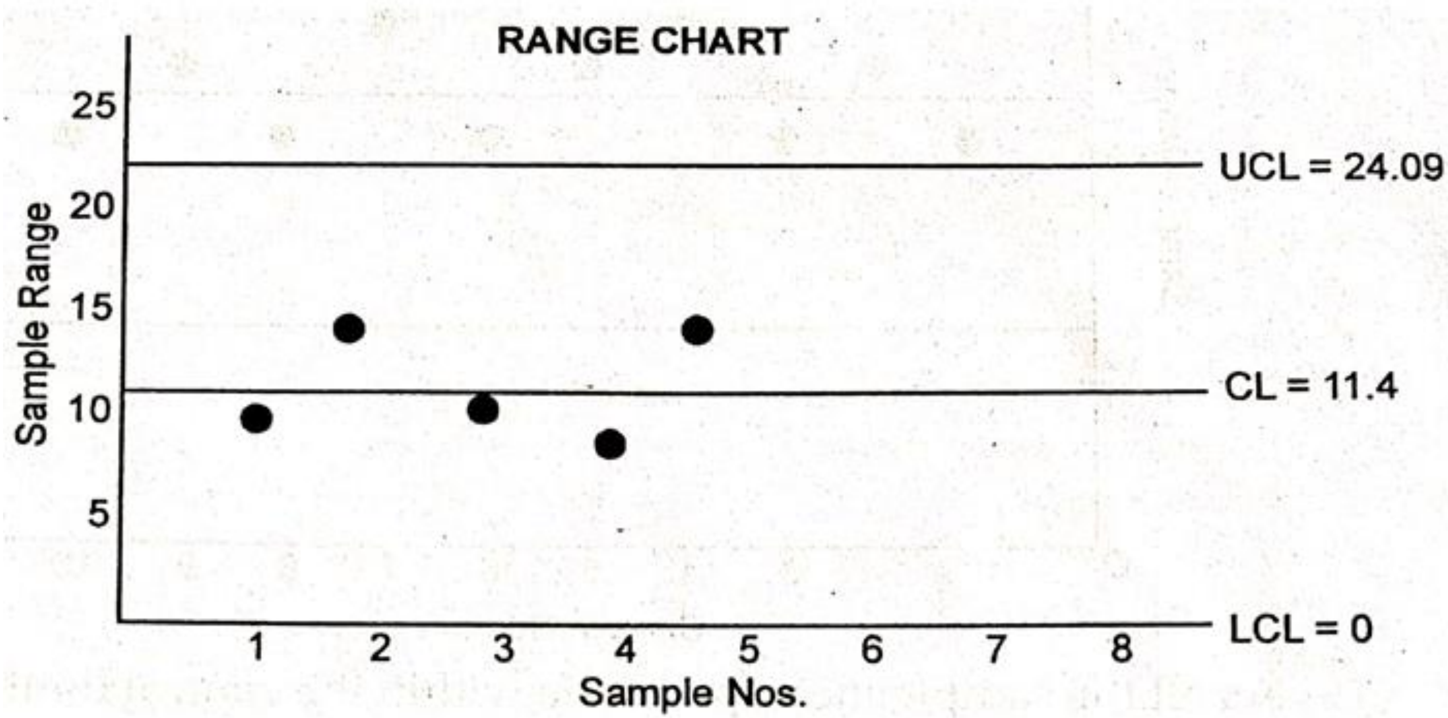
R-chart:

- CL= $\bar{R} = 11.4$

  UCL= $D_4\bar{R} = 24.09$ $(\because D_4 = 2.115)$

  LCL= $D_3\bar{R} = 0$ $(\because D_3 = 0)$

Control Charts (CO2)



As all the range points fall within the control limits, so *R*-chart shows that the given process is in statistical control.

Daily Quiz(CO2)

Q1.An inspection of 10 samples of size 100 each from 10 lots reveal the following number of defectives: 16,18,11,18,21,10,20,18,17,21. Do these indicate that the quality characteristics inspected is under statistical control.

Ans. Process is in under statistical control

NIET Greater Noida

Control Charts (CO2)

Control chart for attributes: are used to monitor characteristic that have discrete values & can be counted. e.g- percentage of defective etc. such charts are of three types-

1. *p*-chart(fraction/proportion defective chart)
2. *np*-chart(no. of defective chart)/d-chart
3. C-chart(no. of defect per unit chart)

Control Charts (CO2)

***p*-chart:** is designed to control the percentage or proportion/fraction of defective per samples.

Construction of *p*-chart:

- Calculate fraction defective p for each sample

i.e. Fraction defective $p = \frac{no.of\ defective\ item}{size\ of\ the\ sample}$

- Calculate avg. fraction defective $(\bar{p})$

i.e. $\bar{p} = \frac{Total\ no.of\ defective}{Total\ no.of\ unit\ inspected}$

Control Charts (CO2)

- CL= $\bar{p}$

$$UCL= \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$LCL= \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where $n = \frac{total\ no.of\ unit\ inspected}{no.of\ sample}$

- If the fraction defective of each sample lie with in the UCL and LCL then process is in control. If it lie outside then process is out of control.

Control Charts (CO2)

Example-1 The following data refers to visual defects found during the inspection of the first 10 samples of size 100 each from a lot of two wheelers manufactured by automobile company:

No. of samples (each of 100 items)	1	2	3	4	5	6	7	8	9	10
No. of Defectives	5	3	3	6	5	6	8	10	10	4

draw from the control chart?

Control Charts (CO2)

Sol: Here no of sample $k = 10$

No.of Samples	No.of units in Sample	No. of defective (d)	Fraction defective $(\mathbf{p})$
1	100	5	0.05
2	100	3	0.05
3	100	3	0.03
4	100	6	0.06
5	100	5	0.05
6	100	6	0.06
7	100	8	0.08
8	100	10	0.1
9	100	10	0.1
10	100	4	0.04
Total	1000	60	0.60

Control Charts (CO2)

Avg. fraction defective $\bar{p} = \frac{Total\ no.of\ defective}{Total\ no.of\ unit\ inspected} = \frac{60}{10\times100} = 0.06$

$$n = \frac{total\ no.of\ unit\ inspected}{no.of\ sample} = \frac{10 \times 100}{10} = 100$$

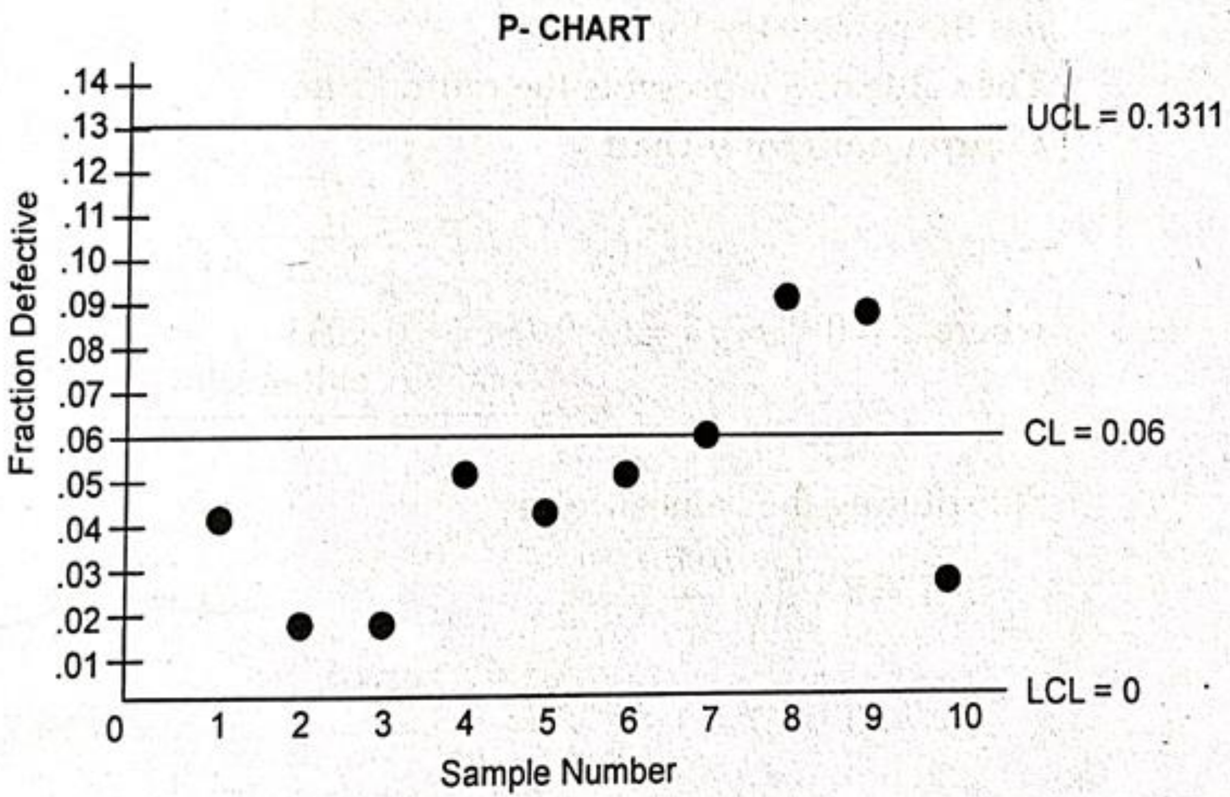
Now CL$= \bar{p} = 0.06$

UCL$= \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.1311$

LCL$= \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = -0.0112$

Here LCL is negative so we take LCL=0.

Control Charts (CO2)



The above chart shows that all the points lie within the control limits. This suggests that the process is in control.

Control Charts (CO2)

***np*-chart:** when the sample size remain constant, *np*- chart is used to control the actual no. of defective per sample. The construction of *np*-chart is similar to *p*- chart because in this chart we can directly plot the no. of defective rather than fraction of defective.

Construction of n*p*-chart:

- Calculate avg. fraction defective ($\bar{p}$)

i.e. $\bar{p} = \frac{Total\ no.of\ defective}{Total\ no.of\ unit\ inspected}$

Control Charts (CO2)

- CL= $n\bar{p}$

UCL= $n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$

LCL= $n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$

where $n = \dfrac{total\ no.of\ unit\ inspected}{no.of\ sample}$

- If the no. of defectives of each sample lie with in the UCL and LCL then process is in control. If it lie outside then process is out of control.

Control Charts (CO2)

Example-1: An inspection of 10 samples of size 400 each from 10 lots reveal the following number of defectives: 17,15,14,26,9,4,19,12,9,15

Calculate the control limits for the number of defective units. Plot on the graph and state whether the process is under control or not.

Sol: Here no. of samples= 10

Size of sample:

$$n = \frac{total\ no.of\ unit\ inspected}{no.of\ sample} = \frac{10 \times 400}{10} = 400$$

Control Charts (CO2)

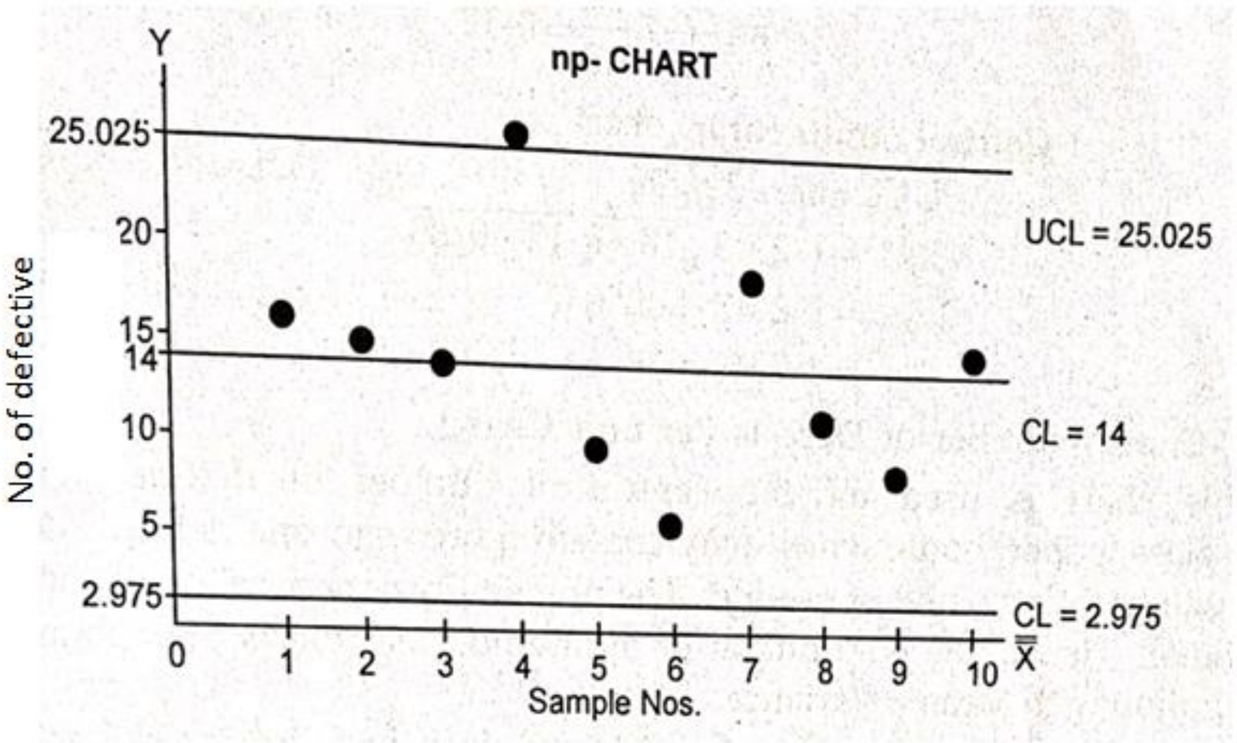
Avg. fraction defective $\bar{p} = \dfrac{Total\ no.of\ defective}{Total\ no.of\ unit\ inspected} = \dfrac{140}{400 \times 10} = 0.035$

CL= $n\bar{p}$ =14

UCL= $n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})} = 25.025$

LCL= $n\bar{p} - 3\ \sqrt{n\bar{p}(1-\bar{p})} = 2.975$

Control Charts (CO2)



Process is out of control.

Control Charts (CO2)

C-chart: is used when the no. of defect per unit are counted. Here C represent the no. of defects per unit (per item)

Construction of C-chart:

- Calculate $\bar{C} = \frac{No.of\ defect\ in\ all\ samples}{total\ no.of\ samples}$
- CL= $\bar{C}$

  UCL= $\bar{C} + 3\sqrt{\bar{C}}$

  LCL= $\bar{C} - 3\ \sqrt{\bar{C}}$

- If the no. of defects with in the UCL and LCL then process is in control. If it lie outside then process is out of control.

Control Charts (CO2)

Example-1 Ten pieces of cloth out of different rolls of equal length contained the following defects: 1,3,5,0,6,0,9,4,4,3

Draw a control chart for the no. of defects and state whether the process is in state of statistical control.

Sol: $\bar{C} = \dfrac{No.of\ defect\ in\ all\ samples}{total\ no.of\ samples} = \dfrac{35}{10} = 3.5$

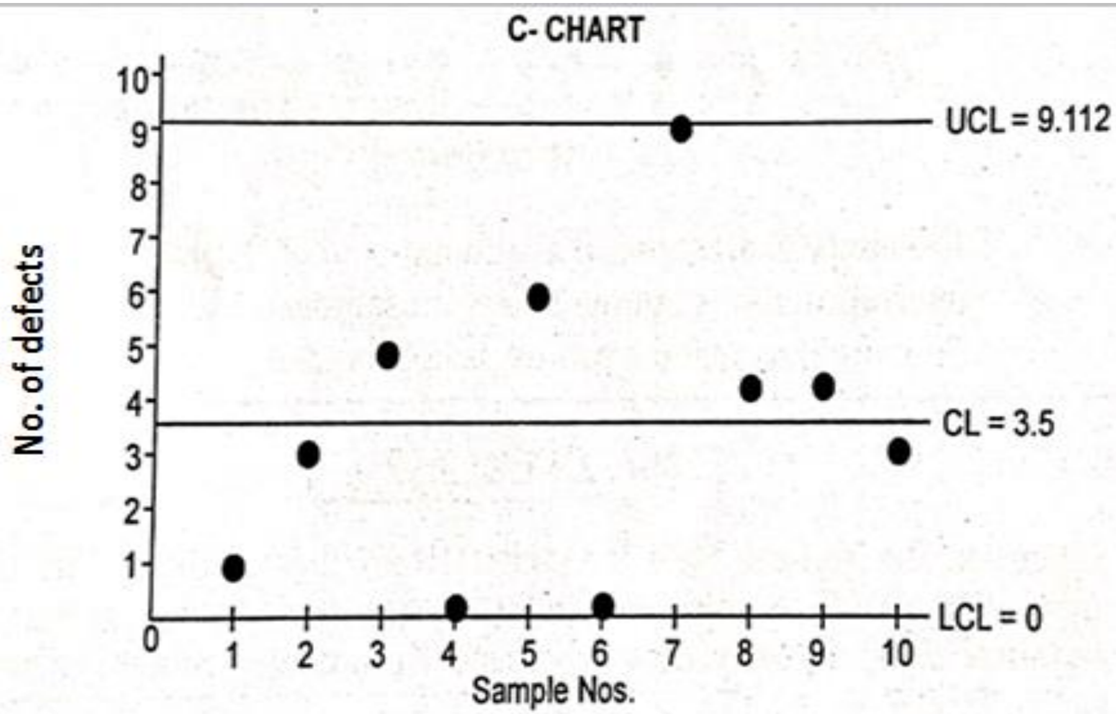
CL= $\bar{C} = 3.5$

UCL= $\bar{C} + 3\sqrt{\bar{C}} = 9.112$

LCL= $\bar{C} - 3\sqrt{\bar{C}} = -2.112$

here LCL is negative so LCL=0

Control Charts (CO2)



The above chart shows that all the plotted points are within the two control limits. This suggests that the process is in control.

Daily Quiz(CO2)

Q1 The no. of defective needles of sewing machines has been given in the following table on the basis of daily inspection:

Day	1	2	3	4	5	6	7	8	9	10
Needle inspected	90	60	70	100	120	50	100	110	100	100
No of defective needle	5	12	7	3	6	5	10	6	8	25

Prepare p-chart and state whether the production process is in control.

Ans: Process is not in control.

Faculty Video Links, Youtube & NPTEL Video Links and Online Courses Details(CO2)

- Self Made Video Link:

Test of hypothesis
https://youtu.be/h-NtACvrPfQ

Test of hypothesis for large samples-Z-test
https://youtu.be/a8MYgSRknmI

Test of hypothesis for large samples-t test
https://youtu.be/msJXJgrY1t4

F-test
https://youtu.be/argjx9DJdxY

Chi square Test
https://youtu.be/ayaa7_e1clA

Faculty Video Links, Youtube & NPTEL Video Links and Online Courses Details(CO2)

Self Made Video Link:

Statistical quality control & control chart for Variable

https://youtu.be/89bs12h_Cg0

Control chart for attributes

https://youtu.be/q6BwAotTI04

Faculty Video Links, Youtube & NPTEL Video Links and Online Courses Details(CO2)

Suggested video link

- Testing of hypothesis
- https://nptel.ac.in/content/storage2/courses/103106120/LectureNotes/Lec3_1.pdf

Weekly Assignment (CO2)

1.The following table gives the yields of four varieties of wheat grown in two plots:

Plots	Varieties			
	A	B	C	D
1	200	230	250	300
2	190	270	300	270
3	240	150	145	180

Is there any significant difference in the production of these varieties?(F value for (8,3)d.f = 4.07 and for (2,6)d.f = 5.14)

Ans. $F_{cal} = 0.385$; Significant

2. The following figures relate to producting in kg. of three varieties A,B and C of wheat shown in 12 plots :

Weekly Assignment (CO2)

A :-		14	16	18	
B :-	14	13	15	22	
C :-	18	16	19	19	20

Is there any significant difference in the production of these varieties?(F value for (2,9)d.f = 4.261)

Ans. $\boldsymbol{F_{cal} = 1.125}$ **; Significant**

3. The number of defects per unit in a Sample of 330 units of a manufactured product was found as follows

No. of defects	**0**	**1**	**2**	**3**	**4**
No. of units	214	92	20	3	1

Weekly Assignment (CO2)

(Given as 5% level of significance, the value of χ^2 for 1 d.f = 3.84).

Ans. $\boldsymbol{\chi^2{}_{cal} = 0.0295}$**; Significant**

4. A milk producer's union wishes to test whether the preference pattern of consumers for its product is dependent on income levels. A random sample of 500 individuals gives the following data:

Income	**Product Preferred**			
	Product A	Product B	Product C	Total
Low	170	30	80	280
Medium	50	25	60	135
High	20	10	55	85
Total	240	65	195	500

Weekly Assignment (CO2)

Can you conclude that the preference patterns are independent of income levels?(Given as 5% level of significance, the value of χ^2 for 4 d.f = 9.49). **Ans.** $\boldsymbol{\chi^2_{cal} = 51.036}$**; Not Significant**

5. Ten pieces of cloth out of different rolls of equal length contained the following number of defects: 1, 3, 5, 0, 6, 0, 9, 4, 4, 3. Draw a control chart for the number of defects and state whether the process is in a state of statistical control.

Ans. Process is in under statistical control

MCQ s (CO2)

1. A sample of 20 items has mean 42 units and S.D. 5 units. It is a random sample from a normal population with mean 45 units. Then in test of hypothesis we get the following results? For this data $t_{tab} = 2.09$ at 5% level of significance.

a) D.f. $\gamma = 19$

b) $|t|_{cal} = 2.615$

c) H_1 accepted

d) There is significant difference between the sample mean and population mean.

2. A random sample of 900 members has mean 3.4cms. If it is reasonably regarded as a sample from a large population of mean 3.2cms and S.D. 2.3 cms. If the tabulated value for this

MCQ s (CO2)

data is 1.96 at 5% level of significance. Then the following results are true.

a) Use t-test for single mean.

b) Use z-test for single mean.

c) Use z-test for double mean.

d) H_0 is accepted.

3.While testing the significance of difference of two sample means in case of small sample, how is degree of freedom calculated?

a) $n_1 - 1$
b) $n_1 + n_2 - 2$
c) $(n_1 - 1) + (n_2 - 1)$
d) $n_2 - 1$

MCQs (CO2)

4. What type of chart will be used to plot the number of defectives in the output of any process?
a) $\bar{X}$ chart
b) R chart
c) c chart
d) p chart

5. Consider a hypothesis H_0 where $\mu = 5$ against H_1 where $\mu > 5$. The test is?
a) Right tailed
b) Left tailed
c) Center tailed
d) Cross tailed

Glossary Questions (CO2)

1. Pick out the correct option from Glossary-

I. z-test

II. F-test

III. t-test

IV. Chi Square test

A. Is used to test the significance for small sample

B. Is used to test the significance for large sample

C. Is also known as variance ratio test

D. Is used to test the independence of attributes

Glossary Questions (CO2)

1. Pick out the correct option from Glossary-

I. np-chart

II. p-chart

III. c-chart

IV. Control chart for variables

A. Fraction defective chart

B. No. of defective chart

C. No. of defect chart

D. Mean chart

Old Question Papers (CO2)

First Sessional Set-1 (CSE,IT,CS,ECE,IOT).docx

Second Sessional Set-2 (CSE,IT,CS,ECE,IOT).docx

Maths IV PUT.docx

Maths IV final paper_2022.pdf

Expected Questions for University Exam (CO2)

1. A random sample of 100 students gave a mean weight of 58 kg with S.D. of 4 kg. Test the hypothesis that mean weight of the population is 60kg. (Take $\alpha = 0.05$)
2. A random sample of 1000 workers from south India shows that their mean wages are Rs.47per week with a S.D. of Rs. 28. A random sample of 1500 workers from north India a mean wages of Rs. 49 per week with a S.D. of Rs. 40. Is there any significant between the mean levels of wages in two places?(Take $\alpha = 0.05$)
3. The 9 items of a sample have the following values: 45,47,50,52,48,47,49,53,51. Does the mean of these values differ significantly from the assumed mean 47.5? (Test at 5% level of significance). (The value of t at 5% level of significance for 8 degree of freedom is 2.31).
4. The height of 6 randomly chosen sailor in inches are 63, 65,68,69,71 and 72. Those of 9 randomly chosen soldiers are 61, 62,65,66,69,70,71,72 and 73. Test at 5% level of significance whether the sailor are on the average taller than soldiers.

Expected Questions for University Exam (CO2)

5. Two independent sample of sizes 7 and 6 have the following values:

Sample A:	28	30	32	33	33	29	34
Sample B:	29	30	30	24	27	29	

Examine whether the samples have been drawn from normal populations having the same variance? (Given that the value of *F* at 5% level of significance for (6,5) d.f. is 4.95 and for (5,6) d.f. is 4.39).

6.Write short notes on following-

i. Null hypothesis and alternate hypothesis

ii. Level of significance

iii. Critical Region

7. A Survey of 240 families with 4 children is given below-

No. of Boys	4	3	2	1	0	Total
No. of Families	10	55	105	58	12	240

Use Chi Square test to state whether the male and female birth are equally possible. The value of $\chi^2$for 4 degree of freedom at 5% level of significance is 9.49.

8. A die is thrown 276 times and the results of these throws are given below:

No. appeared on the die:	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not. (The value of χ^2 at 5% level of significance for 5 degree of freedom is 11.09).

NIET Greater Noida

Recap of unit(CO2)

We discussed the following topics:

- ✓ Test of hypothesis
- ✓ Test of hypothesis for large samples(Z-test)
- ✓ Test of hypothesis for small samples (t-test, F-test)
- ✓ Chi square test One way ANOVA
- ✓ Statistical Quality Control
- ✓ Control charts
- ✓ Control charts for variables(mean and range charts)
- ✓ Control charts for attributes(P chart, np chart, C chart)

References

Text Books

- Erwin Kreyszig, Advanced Engineering Mathematics, 9thEdition, John Wiley & Sons, 2006.
- P. G. Hoel, S. C. Port and C. J. Stone, Introduction to Probability Theory, Universal Book Stall, 2003(Reprint).
- S. Ross: A First Course in Probability, 6th Ed., Pearson Education India, 2002.
- W. Feller, An Introduction to Probability Theory and its Applications, Vol. 1, 3rd Ed., Wiley, 1968.

References

- B.S. Grewal, Higher Engineering Mathematics, Khanna Publishers, 35th Edition, 2000. 2.T.Veerarajan : Engineering Mathematics (for semester III), Tata McGraw-Hill, New Delhi.

- R.K. Jain and S.R.K. Iyenger: Advance Engineering Mathematics; Narosa Publishing House, New Delhi.

- J.N. Kapur: Mathematical Statistics; S. Chand & Sons Company Limited, New Delhi.

- D.N.Elhance,V. Elhance & B.M. Aggarwal: Fundamentals of Statistics; Kitab Mahal Distributers, New Delhi.

Thank You