SMTA1402 - Probability and Statistics

Unit - V Analysis of Variance and Statistical Quality Control

ANOVA (Analysis of Variance) :

Analysis of Variance is a technique that will enable us to test for the significance of the difference among more than two sample means.

Assumptions of analysis of variance:

(i) The sample observations are independent

(ii) The environmental effects are additive in nature

(iii) The samples have been randomly selected from the population.

(iv) Parent population from which observations are taken in normal.

One Way Classification (or) Completely randomized Design (C.R.D)

The C.R.D is the simplest of all the designs, based on principles of randomization and replication. In this design, treatments are allocated at random to the experimental units over the entire experimental materials.

Advantages of completely randomized block design:

The advantages of completely randomized experimental design as follows:

(i) Easy to lay out. (ii) Allow flexibility (iii) Simple statistical analysis

(iv) lots of information due to missing data is smaller than with any other design

Two Way Classification (or) Randomized Block Design (R.B.D):

The entire experiment influences on only two factors is two way Classification.

The basic principles of design of experiments:

(i) Randomization (ii) Replication (iii) Local Control

<u>Working Procedure (One – Way classification)</u>

Null Hypothesis H_0: There is no significance difference between the treatments.

Alternate Hypothesis H_1: There is a significance difference between the treatments.

Analysis:

Step 1: Find N= number of observations

Setp 2: Find T = The total value of observations

Step 3: Find the correction Factor = $C.F = \frac{T^2}{N}$

Step 4: Calculate the total sum of squares = $\text{TSS} = \left(\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + ...\right) - C.\text{F}$

Step 4: Find Total Sum of Square $\text{TSS} = \left(\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + ...\right) - C.\text{F}$

Step 5: Column Sum of Square $\text{SSC}=\left(\frac{\left(\sum X_1\right)^2}{N_1}+\frac{\left(\sum X_2\right)^2}{N_2}+\frac{\left(\sum X_3\right)^2}{N_3}+\ldots\right)-C.\text{F}$

Where $N_i=$ Total number of observation in each column ($i=1,2,3,\ldots$)

Step 6: Prepare the ANOVA TABLE to calculate F-ratio.

Source of Variation	Sum of Degrees	Degree of freedom	Mean Square	F- Ratio
Between Columns	SSC	c-1	$\text{MSC}=\frac{\text{SSC}}{\text{c}-1}$	$\text{F}_\text{C}=\frac{\text{MSC}}{\text{MSE}}$ if $\text{MSC}>\text{MSE}$ (or) $\text{F}_\text{C}=\frac{\text{MSE}}{\text{MSC}}$ if $\text{MSE}>\text{MSC}$
Error	SSE	N-c	$\text{MSE}=\frac{\text{SSE}}{\text{N}-\text{c}}$	
Total				

Step 7: Find the table value (use χ^2 table)

Step 8: Conclusion:

Calculated value < Table Value, the we accept Null Hypothesis H_0 **(or)**

Calculated value > Table Value, the we rejectNull Hypothesis H_0

Working Procedure (Two – Way classification)

Null Hypothesis H_0: There is no significance difference between the treatments.

Alternate Hypothesis H_1: There is a significance difference between the treatments.

Analysis:

Step 1: Find N= number of observations

Setp 2: Find $T=$ The total value of observations

Step 3: Find the correction Factor = $C.F=\frac{T^2}{N}$

Step 4: Calculate the total sum of squares = $\text{TSS}=\left(\sum X_1^2+\sum X_2^2+\sum X_3^2+\ldots\right)-C.\text{F}$

Step 4: Find Total Sum of Square $\text{TSS}=\left(\sum X_1^2+\sum X_2^2+\sum X_3^2+\ldots\right)-C.\text{F}$

Step 5: Find column sum of Square $\text{SSC}=\left(\frac{\left(\sum X_1\right)^2}{N_1}+\frac{\left(\sum X_2\right)^2}{N_2}+\frac{\left(\sum X_3\right)^2}{N_3}+\ldots\right)-C.\text{F}$

Where $N_i=$ Total number of observation in each column $(i=1,2,3,\ldots)$

Step 6: Find Row sum of square = $SSR=\left(\frac{\left(\sum Y_1\right)^2}{N_1}+\frac{\left(\sum Y_2\right)^2}{N_2}+\frac{\left(\sum Y_3\right)^2}{N_3}+...\right)-C.F$

Where N_j = Total number of observation in each Row $(j=1,2,3,...)$

Step 7: Prepare the ANOVA TABLE to calculate F-ratio.

Source of Variation	Sum of Degrees	Degree of freedom	Mean Square	F- Ratio
Between Columns	SSC	c-1	$MSC=\frac{SSC}{c-1}$	$F_C=\frac{MSC}{MSE}$ if $MSC>MSE$ (or) $F_C=\frac{MSE}{MSC}$ if $MSE>MSC$
Between Rows	SSR	r-1	$MSC=\frac{SSR}{r-1}$	$F_R=\frac{MSR}{MSE}$ if $MSR>MSE$ (or) $F_R=\frac{MSE}{MSR}$ if $MSE>MSR$
Error	SSE	N-c-r+1	$MSE=\frac{SSE}{N-c-r+1}$	
Total	TSS	rc-1		

Step 8: Find the table value for both F_C & F_R (use χ^2 table)

Step 9:Conclusion:

Calculated value < Table Value, the we accept Null Hypothesis H_0 **(or)**

Calculated value > Table Value, the we reject Null Hypothesis H_0

1. **The following are the numbers of mistakes made in 5 successive days of 4 technicians working for a photographic laboratory :**

Tech I (X_1)	**Tech II (X_2)**	**Tech III (X_3)**	**Tech IV(X_4)**
6	**14**	**10**	**9**
14	**9**	**12**	**12**
10	**12**	**7**	**8**
8	**10**	**15**	**10**
11	**14**	**11**	**11**

Test at the level of significance α = 0.01 whether the differences among the 4 samples means can be attributed to chance.

Solution:

H_0: There is no significant difference between the technicians

H_1 : Significant difference between the technicians

We shift the origin

	X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
	-4	4	0	-1	**-1**	16	16	0	1
	4	-1	2	2	**7**	16	1	4	4
	0	2	-3	-2	**-3**	0	4	9	4
	-2	0	5	0	**3**	4	0	25	0
	1	4	1	1	**7**	1	16	1	1
Total	**-1**	**9**	**5**	**0**	**13**	**37**	**37**	**39**	**10**

N= Total No of Observations = 20 T=Grand Total = 13

$$\text{Correction Factor} = \frac{(\text{Grand total})^2}{\text{Total No of Observations}} = 8.45$$

$$TSS = \sum X_1{}^2 + \sum X_2{}^2 + \sum X_3{}^2 + \sum X_4{}^2 - C.F = 37 + 37 + 39 + 10 - 8.45 = 114.55$$

$$SSC = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - C.F = \frac{(-1)^2}{5} + \frac{(9)^2}{5} + \frac{(5)^2}{5} + 0 - 8.45 = 12.95$$

SSE = TSS – SSC = 114.55-12.95= 101.6

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Samples	SSC=12.95	C-1= 4-1=3	$MSC = \frac{SSC}{K-1}$ =4.317	$F_C = \frac{MSC}{MSE}$ =1.471
Within Samples	SSE=101.6	N-C=20-4=16	$MSE = \frac{SSE}{N-K}$ =6.35	

Cal F_C = 1.471 & Tab F_C (16,3) = 5.29

Conclusion : Cal F_C < Tab F_C $\Rightarrow$ There is no significance difference between the technicians

2. **A completely randomized design exprement with 10 plots and 3 treatments gave the following results.**

Plot No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Treatment	**A**	**B**	**C**	**A**	**C**	**C**	**A**	**B**	**A**	**B**
Yield	**5**	**4**	**3**	**7**	**5**	**1**	**3**	**4**	**1**	**7**

Analyse the results for treatment effects.

Solution:

A	B	C
5	**4**	**3**
7	**4**	**5**
3	**7**	**1**
1		

Null Hypothesis H_0: There is no significant difference in treatments

Alternate Hypothesis H_1 : Significant difference in treatments

	X_1	X_2	X_3	TOTAL	X_1^2	X_2^2	X_3^2
	5	4	3	**12**	25	16	9
	7	4	5	**16**	49	16	25
Total	3	7	1	**11**	9	49	1
	1			**1**	1		
	16	15	9	**40**	84	81	35

Step1: N= Total No of Observations = 10

Step 2: T=Grand Total = 40

Step 3: Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}} = \dfrac{T^2}{N} = \dfrac{40^2}{10} = 160$

Step 4: TSS= $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 - C.\text{F} = 84 + 81 + 35 - 160 = 40$

Step 5: SSC$= \dfrac{\left(\sum X_1\right)^2}{N_1} + \dfrac{\left(\sum X_2\right)^2}{N_1} + \dfrac{\left(\sum X_3\right)^2}{N_1} - C.\text{F} = \dfrac{(16)^2}{4} + \dfrac{15^2}{3} + 3 - 160$

$SSC = 64 + 75 + 27 - 160 = 6$

Where N_1= Number of elements in each column

Step 7: SSE=TSS-SSC $= 40 - 6 = 34$

Step 8: **ANOVA TABLE:**

Source of Variation	**Sum of Squares**	**Degree of freedom**	**Mean Square**	**F- Ratio**
Between Columns	SSC=6	C-1= 3-1=2	$\text{MSC} = \dfrac{\text{SSC}}{\text{C}-1} = \dfrac{6}{2} = 3$	$\text{F}_\text{C} = \dfrac{\text{MSE}}{\text{MSC}} = \dfrac{4.86}{3} = 1.62$
Error	SSE=34	N-C=10-3=7	$\text{MSE} = \dfrac{\text{SSE}}{\text{N}-\text{C}} = \dfrac{34}{87} = 4.86$	

Cal F_C = 1.62

Table value : $F_C\ (7,2)=19.35$

Conclusion : Cal $F_C <$ Tab F_C

We accept Null Hypothesis $\Rightarrow$ There is no significance difference in tretments

3. **The following table gives the number of articles of a product produced by five different workers using four types of machines.**

Workers	**Machines**			
	P	**Q**	**R**	**S**
A	**44**	**38**	**47**	**36**
B	**46**	**40**	**52**	**43**
C	**34**	**36**	**44**	**32**
D	**43**	**38**	**46**	**33**
E	**38**	**42**	**49**	**39**

Test (i) Whether the five workers differ with respect to mean productivity and
(ii) Whether the four machines differ with respect to mean productivity.

Solution: H_0: There is no significant difference between the Machine types and no significant difference between the Workers

H_1 :Significant difference between the Machine types and no significant difference between the Workers

We shift the origin $X_{ij} = x_{ij} - 46$; $h = 5$; $k = 4$; $N = 20$

	A	B	C	D	Total=T_{i*}	$[T_{i*}^2]/k$	ΣX_{*ij}^2
1	-2	-8	1	-10	**-19**	90.25	169
2	0	-6	6	-3	**-3**	2.25	81
3	-12	-10	-2	-14	**-38**	361	444
4	-3	-8	0	-13	**-24**	144	242
5	-8	-4	3	-7	**-16**	64	138
Total=T_{*j}	**-25**	**-36**	**8**	**-47**	**-100**	661.5	1074
$[T_{*j}^2]/h$	125	259.2	12.8	441.8	838.8		

T=Grand Total = -100

$$\text{Correction Factor} = \frac{(Grand\ \ total)^2}{Total\ \ No\ \ of\ \ Observations} = \frac{(-100)^2}{20} = 500$$

$$TSS = \sum_i \sum_j X_{ij}^2 - C.F = 1074 - 500 = 574$$

$$SSR = \frac{\sum T_{i*}{}^2}{k} - C.F = 661.5 - 500 = 161.5$$

	$SSC = \frac{\sum T_{*j}^{2}}{h} - C.F = 838.8 - 500 = 338.8$ SSE = TSS – SSC – SSR = 574 – 161.5 – 338.8 = 73.7 **ANOVA Table**

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio	F_{Tab}Ratio
Between Rows (Workers)	SSR=161.5	h - 1= 4	MSR= 40.375	F_R = 6.574	$F_{5\%}(4, 12)$ = 3.26
Between Columns (Machine)	SSC=338.8	k – 1=3	MSC = 112.933	F_C = 18.388	$F_{5\%}(3, 12)$ = 3.59
Residual	SSE = 73.7	(h – 1)(k – 1) = 12	MSE = 6.1417		
Total	1074				

Conclusion : Cal F_C< Tab F_C and Cal F_R< Tab F_R $\Rightarrow$ There is no significant difference between the Machine types and no significant difference between the Workers

4. **A Company appointments four salesmen A, B, C and D and observes their sales in 3 seasons: summer, winter and monsoon. The figures (in lakhs of Rs.) are given in the following table:**

Season	**Salesman**			
	A	**B**	**C**	**D**
Summer	**36**	**36**	**21**	**35**
Winter	**28**	**29**	**31**	**32**
Monsoon	**26**	**28**	**29**	**29**

i) Do the salesmen significantly differ in performance?
ii) Is there significant difference between the seasons?
Solution:

Null Hypothesis H_0**:** There is no significant difference between the sales in the 3 seasons and also between the sales of the 4 salesmen.

Alternate Hypothesis H_1**:** There is a significant difference between the sales in the 3 seasons and also between the sales of the 4 salesmen.

Test statistic:
To simplify calculations we deduct 30 from each value

Seasons		A X_1	B X_2	C X_3	D X_4	Seasons Total	X_1^2	X_2^2	X_3^2	X_4^2
Y_1	Summer	6	6	-9	5	**8**	36	36	81	25
Y_2	Winter	-2	-1	1	2	**0**	4	1	1	4
Y_3	Monson	-4	-2	-1	-1	**-8**	16	4	1	1
Total		0	3	-9	6	**0**	56	41	83	30

Step1: N= Total No of Observations = 12

Step 2: T=Grand Total = 0

Step 3: Correction Factor = $\frac{(\text{Grand total})^2}{\text{Total No of Observations}} = \frac{T^2}{N} = \frac{0^2}{12} = 0$

Step 4: TSS= $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 56+41+83+30-0 = 210$

Step 5:

$$\text{SSC} = \frac{\left(\sum X_1\right)^2}{N_1} + \frac{\left(\sum X_2\right)^2}{N_1} + \frac{\left(\sum X_3\right)^2}{N_1} + \frac{\left(\sum X_4\right)^2}{N_1} - C.F = \frac{0^2}{3} + \frac{3^2}{3} + \frac{(-9)^2}{3} + \frac{6^2}{3} - 0$$

$$\text{SSC} = 0+3+27+12-0 = 42$$

Where N_1= Number of elements in each column

Step 6:

$$\text{SSR} = \frac{\left(\sum Y_1\right)^2}{N_2} + \frac{\left(\sum Y_2\right)^2}{N_2} + \frac{\left(\sum Y_3\right)^2}{N_2} - C.F = \frac{8^2}{4} + \frac{0^2}{4} + \frac{(-8)^2}{4} + \frac{6^2}{4} - 0 = 16+0+16-0 = 32$$

Where N_2= Number of elements in each row

Step 7: SSE=TSS-SSC-SSR $= 210-42-32$

Step 8: **ANOVA TABLE:**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	varience	F – ratio
Between Columns (Salesmen)	SSC=42	c-1=4-1=3	$MSC = \frac{SSC}{c-1} = \frac{42}{3} = 14$	$MSC = \frac{MSE}{MSC} = \frac{22.67}{14} = 1.619$	$F_C(6,3) = 8.94$
Between rows (Seasons)	SSR =32	r-1=3-1=2	$MSR = \frac{SSR}{r-1} = \frac{32}{2} = 16$	$MSR = \frac{MSE}{MSR} = \frac{22.67}{16} = 1.417$	$F_R(6,2) = 8.94$

Error	SSE=136	N-c-r +1=6	$MSE = \frac{SSE}{N-c-r+1} = \frac{136}{6} = 22.67$		
Total	210	11			

Table Value of F = $F_C(\text{Error}, \text{d.f}) = F_C(6,3) = 8.94$, $F_R(\text{Error}, \text{d.f}) = 8.94$ with 5% level of significance

Conclusion:

1) Cal F_R **< Table** $F_{R,0.05}(6,3)$

Hence we accept the H_0 and we conclude that there is no significant difference between sales in the three seasons.

2) **Cal** F_R **< Table** $F_{R,0.05}(6,2)$.

Hence we accept the H_0 and we conclude that there is no significant difference between in the sales of 4 salesmen.

5. **Analyze 2^2 factorial experiments for the following table.**

Treatment	Replications			
	I	II	III	IV
(1)	12	12.3	11.8	11.6
a	12.8	12.6	13.7	14
b	11.5	11.9	12.6	11.8
ab	14.2	14.5	14.4	15

SOLUTION:

Null hypothesis: All the mean effects are equal.

Let A and B be the two factors.

Let n=number of replications=4

Subtract 12 from each

Treatment	Replications			
	I	II	III	IV
(1)	0	0.3	-0.2	-0.4
a	0.8	0.6	1.7	2
b	-0.5	-0.1	0.6	-0.2
ab	2.2	2.5	2.4	3

Let us find SS for the table

Treatment	**Replications**				**Row Total R_i**	R_i^2
	I	**II**	**III**	**IV**		
(1)	**0**	**0.3**	**-0.2**	**-0.4**	**-0.3**	**0.09**
a	**0.8**	**0.6**	**1.7**	**2**	**5.1**	**26.01**
b	**-0.5**	**-0.1**	**0.6**	**-0.2**	**-0.2**	**0.04**
ab	**2.2**	**2.5**	**2.4**	**3**	**10.1**	**102.01**
Column Total C_j	**2.5**	**3.3**	**4.5**	**4.4**	**T=14.7**	
C_j^2	**6.25**	**10.89**	**20.25**	**19.36**		

T=14.7

Correction factor= $\frac{T^2}{N}$ **=13.5**

TSS=21.19

SSC=0.688

SSR=18.54

SSE=1.962

Source of Variation	**Sum of Squares**	**Degree of freedom**	**Mean Square**	**F- Ratio**	**F_{Tab} Ratio**
b	S_B=1.63	1	MSB=1.63	$F_B = 7.409$	10.56
a	S_A=15.41	1	MSA=15.41	$F_A = 70.04$	10.56

ab	S_{AB}=1.50	1	MSAB=1.50	$F_{AB}=6.81$	10.56
Error	SSE=1.962	N-C-r+1=9	SSE=1.962		

Cal(F_A)=70.04 $\Rightarrow H_0$ is rejected at 1% level

Cal(F_B)=7.409 $\Rightarrow H_0$ is accepted at 1% level

Cal(F_{AB})=10.56 $\Rightarrow H_0$ is accepted at 1% level

6. **Analyse the variance in the following latin square of yields (in kgs) of paddy where A, B, C, D denote the different methods of cultivation.**

D	**122**	**A**	**121**	**C**	**123**	**B**	**122**
B	**124**	**C**	**123**	**A**	**122**	**D**	**125**
A	**120**	**B**	**119**	**D**	**120**	**C**	**121**
C	**122**	**D**	**123**	**B**	**121**	**A**	**122**

Examine whether the different methods of cultivation have given significantly different yields.

Solution:

We shift the origin $X_{ij} = x_{ij} - 100$; n = 4; N = 16

	I	**II**	**III**	**IV**	**Total=T_{i*}**	**$[T_{i*}^2]/n$**	**ΣX_{*ij}^2**
A	2	1	3	2	**8**	16	18
B	4	3	2	5	**14**	49	54
C	0	-1	0	1	**0**	0	2
D	2	3	1	2	**8**	16	18
Total=T_{*j}	**8**	**6**	**6**	**10**	**30**	81	92
$[T_{*j}^2]/n$	16	9	9	25	59		
ΣX_{i*}^2	24	20	14	34	92		

	Letters				Total=T_{i*}	$[T_{i*}^2]/n$
P	1	2	0	2	**5**	6.25
Q	2	4	-1	1	**6**	9
R	3	3	1	2	**9**	20.25
S	2	5	0	3	**10**	25
Total					**30**	60.5

T=Grand Total = 30 ;Correction Factor = $\dfrac{(Grand\ \ total)^2}{Total\ \ No\ \ of\ \ Observations} = \dfrac{(30)^2}{16}$

$$TSS = \sum_i \sum_j X_{ij}^2 - C.F = 92 - \frac{(30)^2}{16} = 35.75$$

$$SSR = \frac{\sum T_{i*}^{\ 2}}{n} - C.F = 81 - \frac{(30)^2}{16} = 24.75$$

$$SSC = \frac{\sum T_{*j}^{\ 2}}{n} - C.F = 59 - \frac{(30)^2}{16} = 2.75$$

$$SSL = \frac{\sum T_{i*}^{\ 2}}{n} - C.F = 60.5 - \frac{(30)^2}{16} = 4.25$$

SSE = TSS – SSC – SSR-SSL = 35.75 – 24.75 – 2.75 – 4.25 = 4

ANOVA Table

Source of Variation	**Sum of Squares**	**Degree of freedom**	**Mean Square**	**F- Ratio**	**F_{Tab}Ratio (5% level)**
Between Rows	SSR=24.75	n - 1= 3	MSR=8.25	F_R= 12.31	F_R(3, 6)=4.76
Between Columns	SSC=2.75	n - 1= 3	MSC = 0.92		
Between Letters	SSL = 4.25	n - 1= 3	MSL = 1.42	F_C = 1.37	Fc(3, 6)=4 .76
Residual	SSE= 4	(n – 1)(n – 2) = 6	MSE = 0.67	F_L = 2.12	

					$F_L(3, 6)$=4 .76
Total	35.75				

Conclusion :
Cal F_C< Tab F_C , Cal F_L< Tab F_L and Cal F_R> Tab F_R $\Rightarrow$ There is significant difference between the **rows** , no significant difference between the letters and no significant difference between the columns

7. **A variable trial was conducted on wheat with 4 varieties in a Latin Square Design. The plan of the experiment and the per plot yield are given below :**

C	**25**	**B**	**23**	**A**	**20**	**D**	**20**
A	**19**	**D**	**19**	**C**	**21**	**B**	**18**
B	**19**	**A**	**14**	**D**	**17**	**C**	**20**
D	**17**	**C**	**20**	**B**	**21**	**A**	**15**

Analyse data and interpret the result.

H_0 : Four varieties are similar

H_1 : Four varieties are not similar

Let us take 20 as origin for simplifying the calculation

Variety	**X_1**	**X_2**	**X_3**	**X_4**	**TOTAL**	**X_1^2**	**X_2^2**	**X_3^2**	**X_4^2**
Y_1	5	3	0	0	**8**	25	9	0	0
Y_2	-1	-1	1	-2	**-3**	1	1	1	4
Y_3	-1	-6	-3	0	**-10**	1	36	9	0
Y_4	-3	0	1	-5	**-7**	1	0	1	25
	0	**-4**	**-1**	**-7**	**-12**	**9**	**46**	**11**	**29**

N=Total No of Observations = 16 T=Grand Total = -12

$$\text{Correction Factor} = \frac{(\text{Grand total})^2}{\text{Total No of Observations}} = 9$$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 36 + 46 + 11 + 29 - 9 = 113$$

$$SSC = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - C.F = \frac{(6)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 9 = 4$$

$$SSR = \frac{(\sum Y_1)^2}{N_1} + \frac{(\sum Y_2)^2}{N_2} + \frac{(\sum Y_3)^2}{N_2} + \frac{(\sum Y_4)^2}{N_2} - C.F = \frac{(8)^2}{4} + \frac{(-3)^2}{4} + \frac{(-10)^2}{4} + \frac{(-7)^2}{4} - 9 = 46.5$$

To find SSK

Treatment	1	2	3	4	Total
A	0	-1	-6	-5	-12
B	3	-2	-1	1	1
C	5	1	0	0	6
D	0	-1	-3	-3	-7

SSK= $\frac{(\sum Y_1)^2}{K_1} + \frac{(\sum Y_2)^2}{K_2} + \frac{(\sum Y_3)^2}{K_3} + \frac{(\sum Y_4)^2}{K_4} - C.F$

$= 48.5$

SSE= TSS − SSC−SSR−SSK = 113-7.5-46.5-48.5=10.5

ANOVA Table

Source of Variation	**Sum of Squares**	**Degree of freedom**	**Mean Square**	**F- Ratio**
Column Treatment	SSC=7.5	n-1=3	$MSC = \frac{SSC}{n-1}$ =2.5	$F_C = \frac{MSC}{MSE}$=1.43
Row Treatments	SSR=46.5	n-1=3	$MSR = \frac{SSR}{n-1}$ =15.5	$F_R = \frac{MSE}{MSR}$=8.86
Between Treatments	SST=48.5	n-1=3	$MSK = \frac{SSK}{n-1}$ =16.17	$F_K = \frac{MSK}{MSE}$=9.24
Error (or) Residual	SSE=10.5	(n-1) (n-2)=6	$MSE = \frac{SSE}{(n-1)(n-2)} = 1.75$	

Table value F(3,6) degrees of freedom 8.94

There is significant difference between treatments

STATISTICAL QUALITY CONTROL

Statistical quality control:

statistical quality control is a statistical method for finding whether the variation in the quality of the product is due to random causes or assignable causes.

Objectives of statistical quality control:

To achieve better utilization of raw materials, to control waste and scrap and to optimize the quality of the product without any defects.

Control chart:

It is a useful graphical method to find whether a process is in statistical quality control.

Uses of Quality control chart:

It helps in determining whether the goal set is being achieved by finding out whether the Process is in control or not.

Different types of control chart:

Control chart for variables – Range and mean chart, Control chart for attributes- p-chart,

C-chart, np-chart.

control limits for mean chart:

Central limit = $\bar{\bar{x}}$ **, upper control limit =** $\bar{\bar{x}} + A_2\bar{R}$ **, lower control limit =** $\bar{\bar{x}} - A_2\bar{R}$

Where $\bar{x}$ **is the mean of the sample and R is the range.**

The control limits for range chart:

CL= $\bar{R}$ **, UCL =** $D_4\bar{R}$ **, LCL=**$D_3\bar{R}$ **.**

Procedure to draw the $\bar{x}-$chart **& R-chart:**

1. The sample values in each of the N samples each of size 'n' will be given. Let $\overline{X_1}, \overline{X_2}, \ldots \overline{X_N}$ **be the means of the N samples &** $R_1, R_2 \ldots R_N$ **be the ranges of the N samples.**

2. Compute $\overline{\overline{X}} = \frac{1}{N}\left(\overline{X_1} + \overline{X_2} + \ldots + \overline{X_N}\right); \overline{R} = \frac{1}{N}\left(R_1 + R_2 + \ldots + R_N\right)$

3. The values of A_2, D_3, D_4 **for the given sample size n are taken from the table of control chart constants.**

4. Find the values of the control limits $\bar{\bar{x}} \pm A_2\bar{R}$ **(for the mean chart) and the control limits** $D_3\bar{R}$ **and** $D_4\bar{R}$ **(for the range chart) are computed.**

5. On the ordinary graph sheet, the sample numbers are represented on the x-axis and the sample means on the

y-axis (for the mean chart) and the sample ranges on the y-axis(for the range chart).

6.For drawing the mean chart, we draw the three lines $y = \overline{\overline{X}}$ **,** $y = \overline{\overline{X}} - A_2\overline{R}$ **and** $y = \overline{\overline{X}} + A_2\overline{R}$ **which represent respectively the central line, the L.C.L line and U.C.L line, Also we plot the points whose coordinates are** $(1, \overline{X_1}),(2, \overline{X_2}),\ldots(N, \overline{X_N})$ **and join adjacent points by line segments. The graph thus obtained is the** $\overline{X}$ **chart.**

7. For drawing the mean chart, we draw the three lines $y = \overline{R}$**,** $y = D_3\overline{R}$ **and** $y = D_4\overline{R}$ **which represent respectively the central line, the L.C.L line and U.C.L line, Also we plot the points whose coordinates are** $(1, R_1),(2, R_2),\ldots(N, R_N)$ **and join adjacent points by line segments. The graph thus obtained is the** R **chart.**

Mean And Range Chart Problems

1. **Given below are the values of sample mean $\overline{X}$ and sample range R for 10 samples, each of size 5. Draw the appropriate mean and range charts and comment on the state of control on the state of control of the process.**

Sample No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean $\overline{X}_i$	**43**	**49**	**37**	**44**	**45**	**37**	**51**	**46**	**43**	**47**
Range R_i	**5**	**6**	**5**	**7**	**7**	**4**	**8**	**6**	**4**	**6**

Solution:

$$\overline{\overline{X}} = \frac{1}{N}\sum \overline{X_i}$$
$$= \frac{1}{10}[43 + 49 + 37 + 44 + 45 + 37 + 51 + 46 + 43 + 47]$$
$$= 44.2$$

$$\overline{R} = \frac{1}{N}\sum R_i$$
$$= \frac{1}{10}[5 + 6 + 5 + 7 + 7 + 4 + 8 + 6 + 4 + 6]$$
$$= 5.8$$

From the table of control chart for sample size n=5, we have $A_2 = 0.577, D_3 = 0 \,\&\, D_4 = 2.115$

i) **Control limits for $\overline{X}$ chart:**

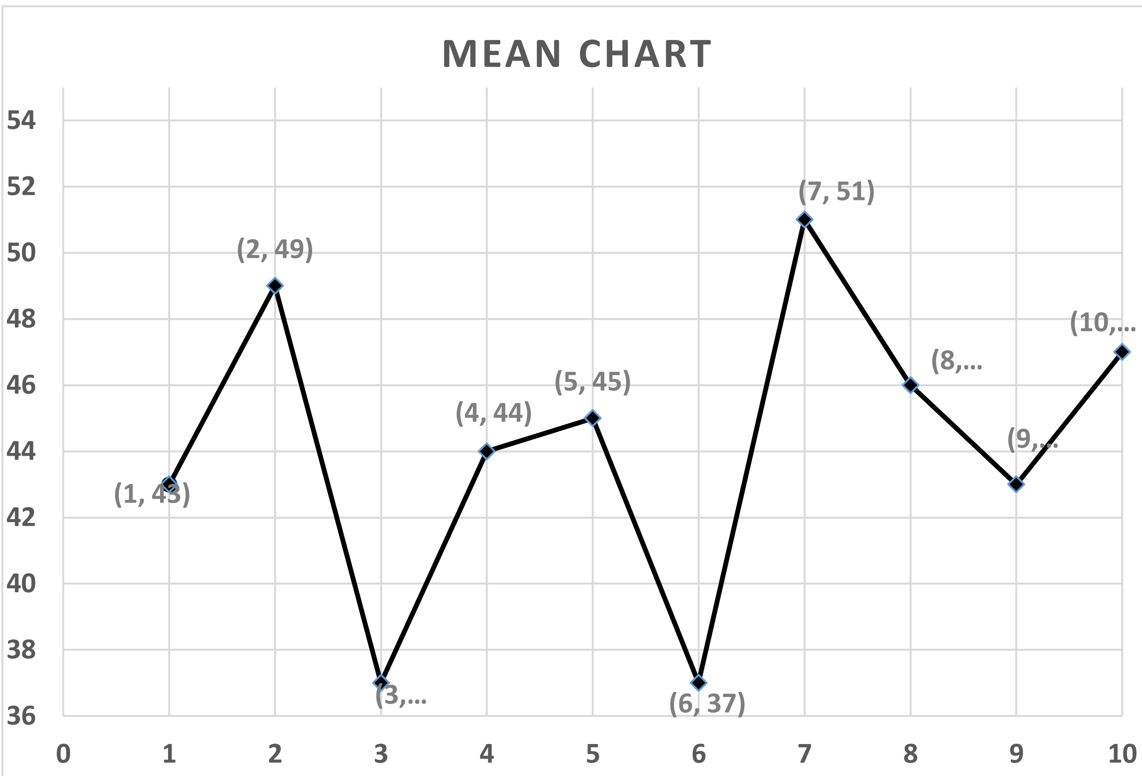
CL (central line) = $\overline{\overline{X}} = 44.2$

$$LCL = \overline{\overline{X}} - A_2\overline{R} = 44.2 - (0.577)(5.8) = 40.85$$

$$UCL = \overline{\overline{X}} + A_2\overline{R} = 44.2 + (0.577)(5.8) = 47.55$$

Conclusion :

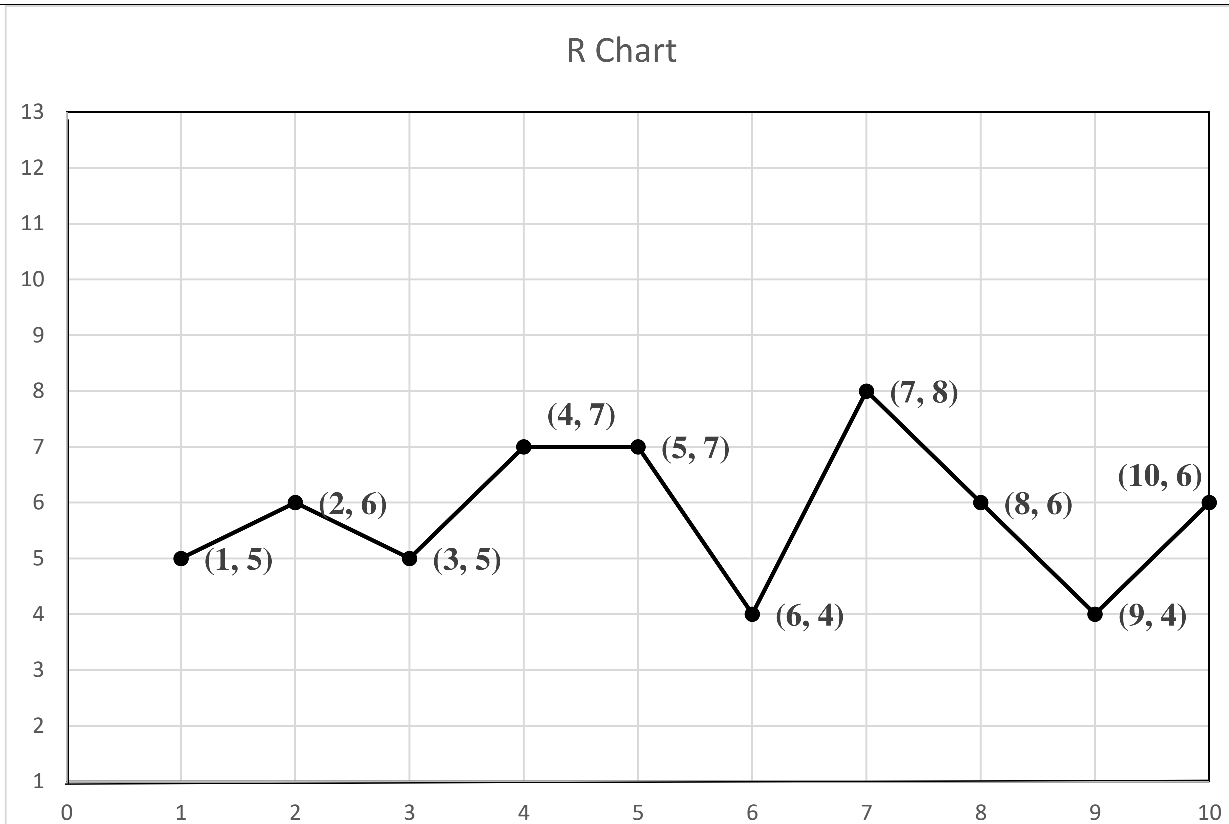
Since 2^{nd}, 3^{rd}, 6^{th} and 7^{th} sample means fall outside the control limits the statistical process is out of control according to $\overline{X}$ *chart*



ii) Control limits for R-Chart:

$$CL = \overline{R} = 5.8; \quad LCL = D_3; \quad \overline{R} = 0$$

$$LCL = D_4\overline{R} = (2.115)(5.8) = 12.267 \approx 12.27$$



Conclusion :

Since all the sample means fall within the control limits the statistical process is under control according to $R\,chart$.

2. **The following data give the measurements of 10 samples each of size 5 in the production process taken in an interval of 2 hours. Calculate the sample means and ranges and draw the control charts for mean and range.**

Sample No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Observed measurements X	**49**	**50**	**50**	**48**	**47**	**52**	**49**	**55**	**53**	**54**
	55	**51**	**53**	**53**	**49**	**55**	**49**	**55**	**50**	**54**
	54	**53**	**48**	**51**	**50**	**47**	**49**	**50**	**54**	**52**
	49	**46**	**52**	**50**	**44**	**56**	**53**	**53**	**47**	**54**
	53	**50**	**47**	**53**	**45**	**50**	**45**	**57**	**51**	**56**

Solution:

$$\overline{\overline{X}} = \frac{1}{N}\sum \overline{X_i}$$

$$= \frac{1}{10}\left[52+50+50+51+47+52+49+54+51+54\right]$$

$$= 51.0$$

4

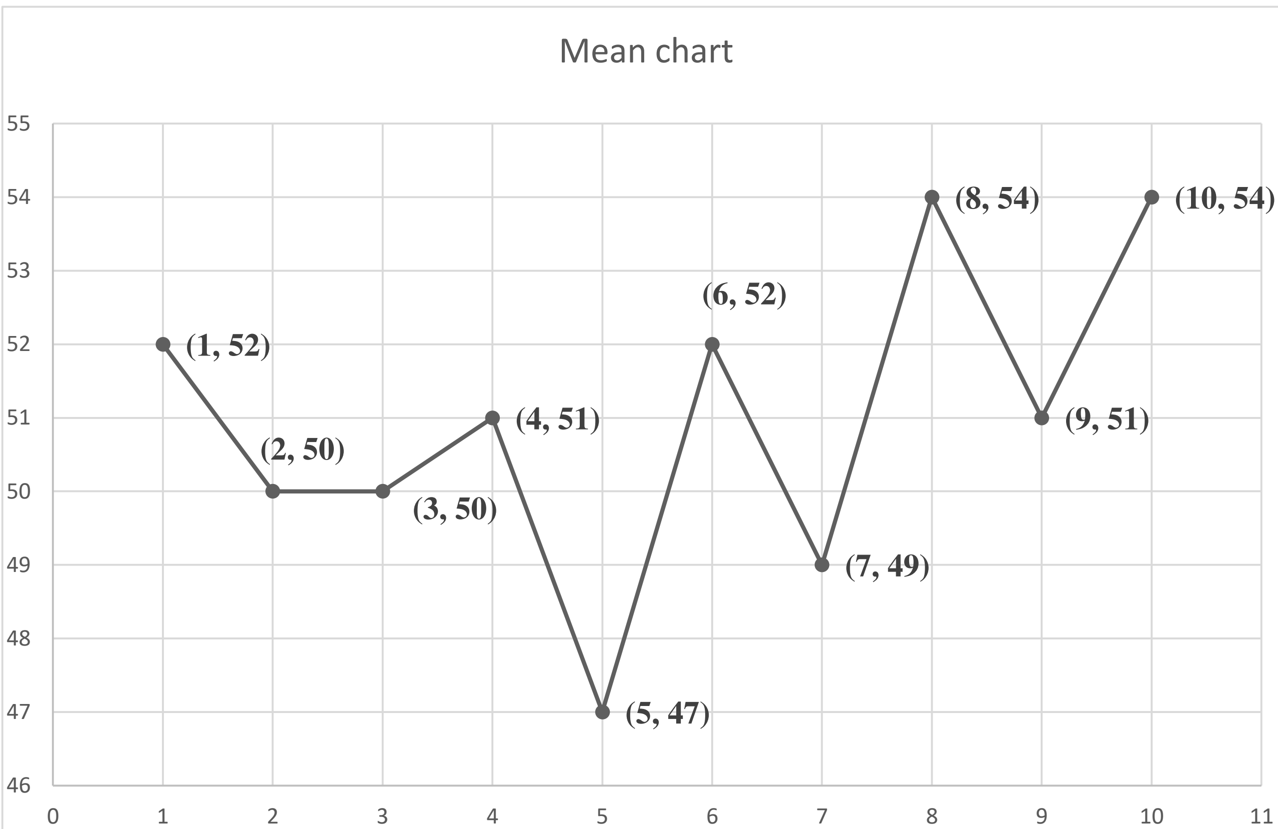
$$\overline{R} = \frac{1}{N}\sum R_i$$
$$= \frac{1}{10}[6+7+6+5+6+9+8+7+7+4]$$
$$= 6.5$$

From the table of control chart for sample size n=5, we have $A_2 = 0.577, D_3 = 0 \,\&\, D_4 = 2.115$

i) **Control limits for $\overline{X}$ chart:**

CL (central line) = $\overline{\overline{X}} = 44.2$

$$LCL = \overline{\overline{X}} - A_2\overline{R}_2 = 51.0 - (0.577)(6.5) = 47.2495$$
$$UCL = \overline{\overline{X}} + A_2\overline{R}_2 = 51.0 + (0.577)(6.5) = 54.7505$$
$$CL = \overline{X} = 51.0$$

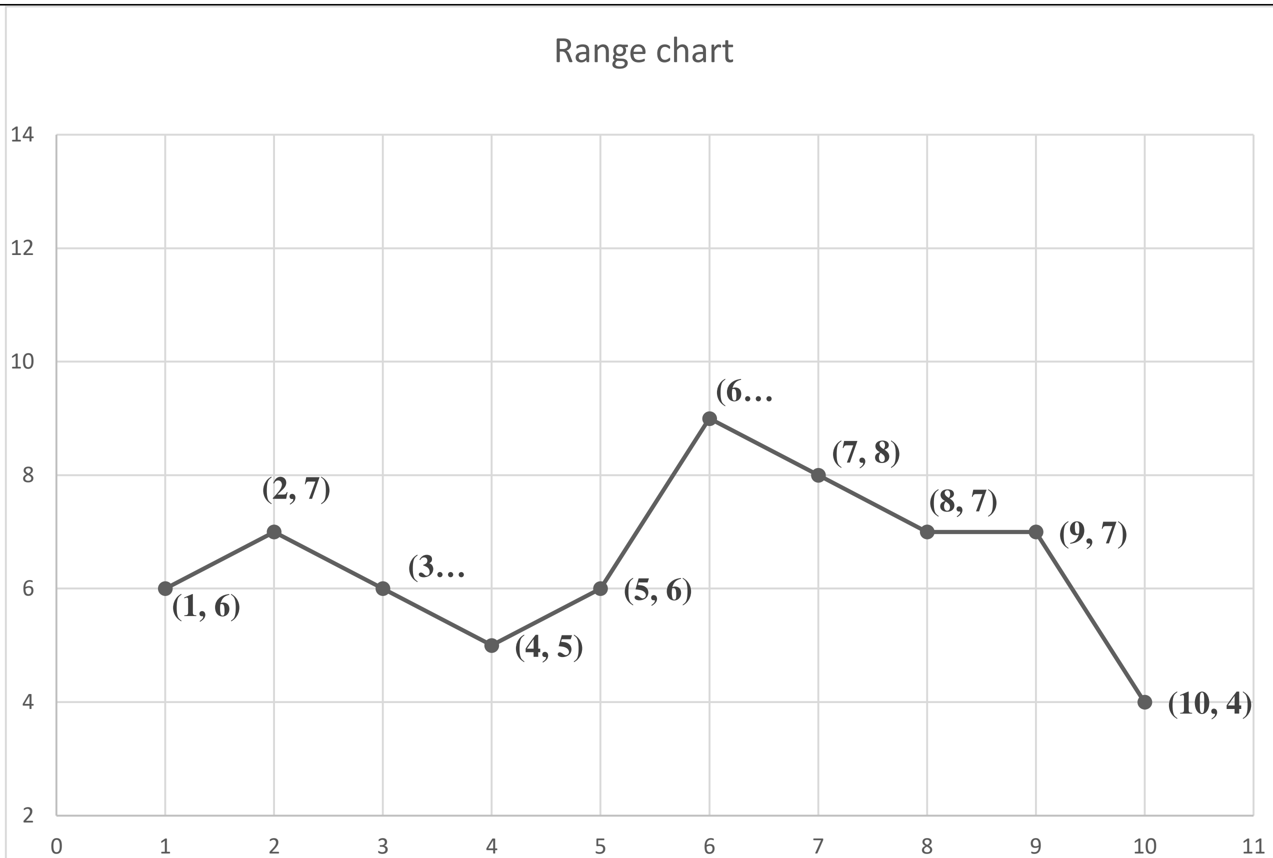


Conclusion :

Since 5th sample mean fall outside the control limits the statistical process is out of control according to $\overline{X}$ *chart*

ii) **Control limits for R-Chart:**

$$CL = \overline{R} = 6.5; \quad LCL = D_3\ \overline{R} = 0$$
$$UCL = D_4\overline{R} = (2.115)(6.5) = 13.7475$$



Conclusion :

Since all the sample means fall within the control limits the statistical process is under control according to $R\,chart$.

C-chart:

Control chart for number of defects is called c-chart.

The control limits for c-chart.

CL = $\bar{c}$ **UCL =** $\bar{c} + 3\sqrt{c}$ **LCL =** $\bar{c} - 3\sqrt{c}$

C-Chart problems

1. **15 tape recorders were examined for quality control test. The number of defects in each tape recorder is recorded below. Draw the appropriate control chart and comment on the state of control.**

Unit No.(i)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of defects (c)	2	4	3	1	1	2	5	3	6	7	3	1	4	2	1

Solution:

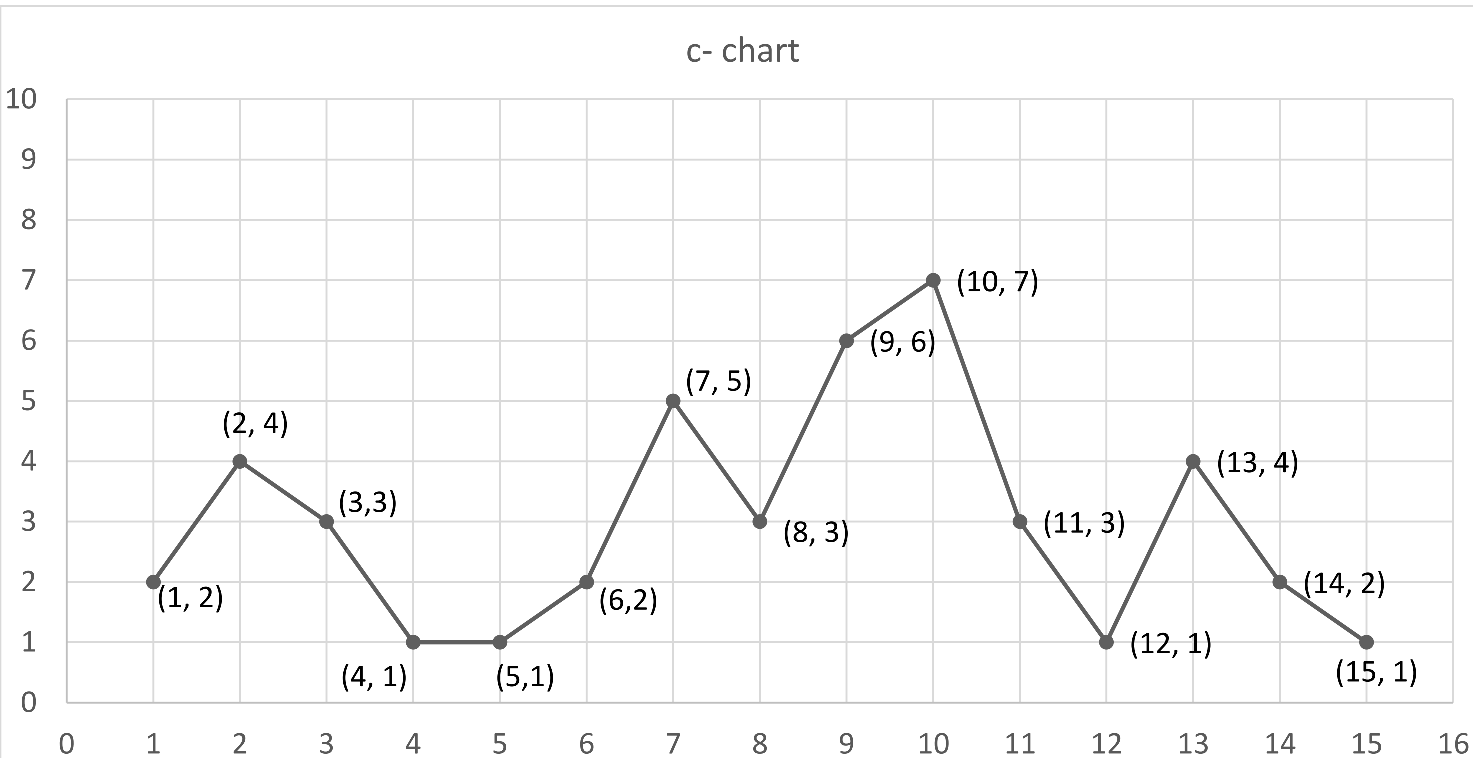
The number of defects per sample containing only one item is given,

$$\bar{c} = \frac{\sum c_i}{N} = \frac{(2+4+3+\cdots+2+1)}{15} = \frac{45}{15} = 3$$

$\mathrm{CL} = \bar{\mathrm{c}} = 3;\quad \mathrm{LCL} = \bar{\mathrm{c}} - 3\sqrt{\mathrm{c}} = 3 - 3\sqrt{3} = -2.20$

We take $\mathrm{LCL} = 0$ (since LCL cannot be –ve)

$$\mathrm{UCL} = \bar{\mathrm{c}} + 3\sqrt{\mathrm{c}} = 3 + 3\sqrt{3} = 8.20$$



Since all the sample points lie within the LCL and UCL lines, the process is under control.

2. 20 pieces of cloth out of different rolls contained respectively 1,4,3,2,4,5,6,7,2,3,2,5,7,6,4,5,2,1,3 and 8

imperfections.

Ascertain whether the process is in a state of statistical control.

Solution:

Let C denote the number of imperfections per unit.

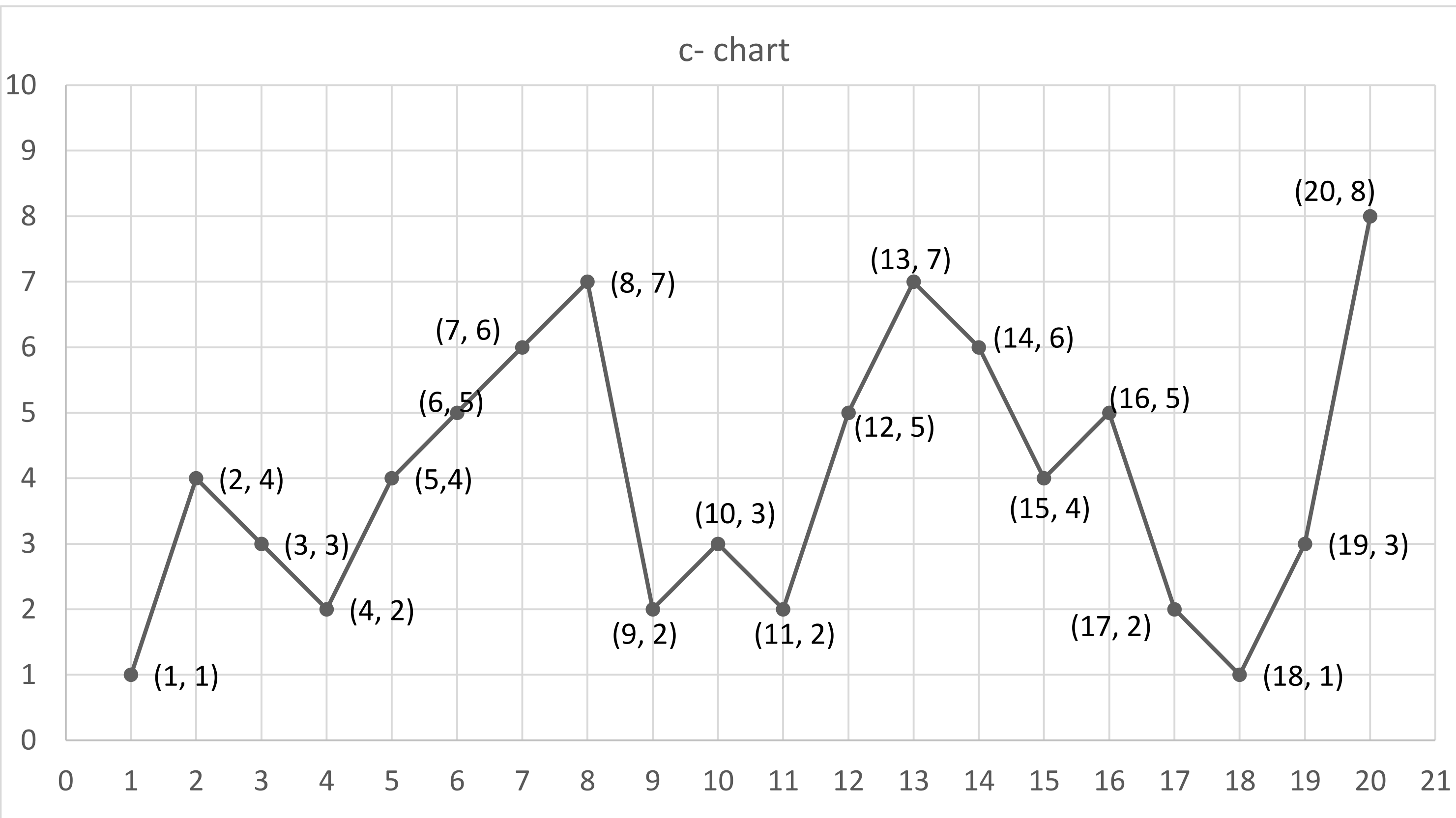
$$\bar{c} = \frac{Total\ no\ of\ defects}{Total\ sample\ inspected} = \frac{\sum c}{n}$$

$$\bar{c} = \frac{1+4+3+2+4+5+\ldots+1+3+8}{20} = \mathbf{4}$$

UCL= $\bar{C} + 3\sqrt{\bar{C}} = 10$

LCL= $\bar{C} - 3\sqrt{\bar{C}} = -2$

We take $\text{LCL} = 0$ (since LCL cannot be –ve)



Since all the sample points lie within the LCL and UCL lines, the process is under control.

p-chart:

Control chart for fraction defectives is called p-chart.

control limits for p-chart.

UCL= $\mathbf{n\bar{p} + 3\sqrt{n\bar{p}\bar{q}}}$, **LCL=**$\mathbf{n\bar{p} - 3\sqrt{n\bar{p}\bar{q}}}$ **CL =** $\mathbf{n\bar{p}}$

np –chart.

Control chart for number of defectives is called np chart.

P-Chart & nP-Chart Problems

1. **Construct a control chart for defectives for the following data:**

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
No. inspected :	**90**	**65**	**85**	**70**	**80**	**80**	**70**	**95**	**90**	**75**
No. of defectives:	**9**	**7**	**3**	**2**	**9**	**5**	**3**	**9**	**6**	**7**

Solution:

We note that the size of the simple varies from sample to sample. We can construct P-chart, provided 0.75 $\bar{n} < n_i$ $<1.25\bar{n}$, for all i.

Here

$$\bar{n} = \frac{1}{N}\sum n_i = \frac{1}{10}(90+65+.....+90+75)$$
$$= \frac{1}{10}(800) = 80$$

Hence The values of n_i be between 60 and 100. Hence p-chart can be drawn by the method given below. Now $\bar{p}$ = $\frac{Total\ no.of\ defectives}{Total\ no.of\ items\ inspected}$

$$= \frac{60}{800} = 0.075$$

Hence for the p-chart to be constructed, CL= $\bar{p}$ =0.075

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

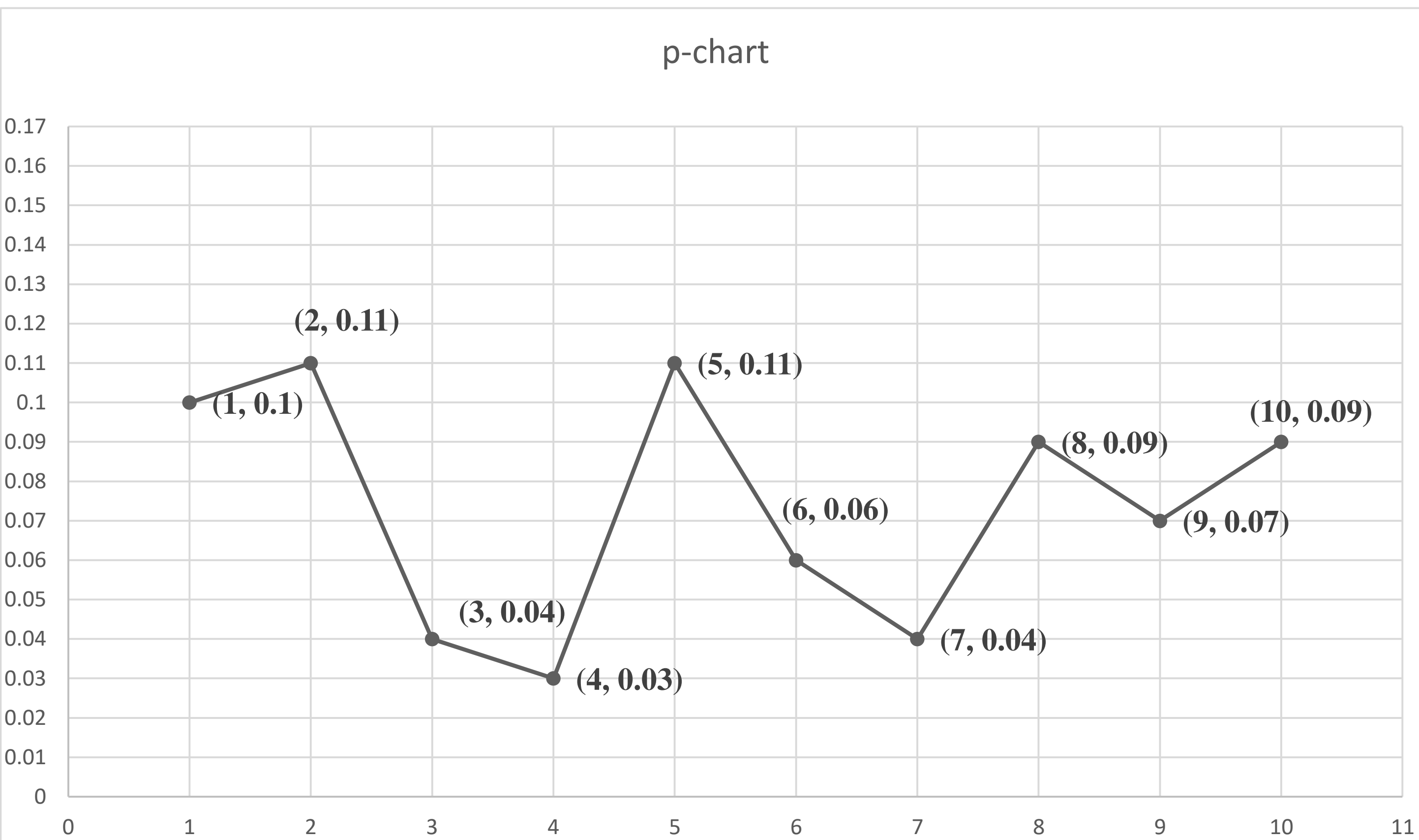
$$=0.075-3\sqrt{\frac{0.075\,X\,0.925}{80}}=-0.013$$

Since LCL cannot be negative,it is taken 0.

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

$$=0.075+3\sqrt{\frac{0.075\,X\,0.925}{80}}=0.163$$

The values of p_i for the various samples are 0.100, 0.108, 0.035, 0.029, 0.113, 0.063, 0.043, 0.095, 0.067, 0.093



Since all the sample points lie within the control lines, the process is under control.

2. **The following are the figures for the number of defectives of 10 samples each containing 100 items 8,10,9,8,10,11,7,9,6,12 .Draw control chart for fraction defective and comment on the state of control of the process.**

P for sample = $= \dfrac{No.of\ defectives\, in\, the\, sample}{No.of\ items\, in\, the\, sample}$

P for sample = $\dfrac{8}{100} = 0.08$

Similarly calculate p for each sample and tabulate.Divide the number of defectives by 100 to get the fraction defective.

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
No. of defectives:	**8**	**10**	**9**	**8**	**10**	**11**	**7**	**9**	**6**	**12**
P=fraction defectives	**0.08**	**0.10**	**0.09**	**0.08**	**0.10**	**0.11**	**0.07**	**0.09**	**0.06**	**0.12**

$$\bar{p} = \frac{\sum p}{n}$$

$$= \frac{0.08 + 0.10 + 0.09 + 0.08 + 0.10 + 0.11 + 0.07 + 0.09 + 0.06 + 0.12}{10} = 0.09$$

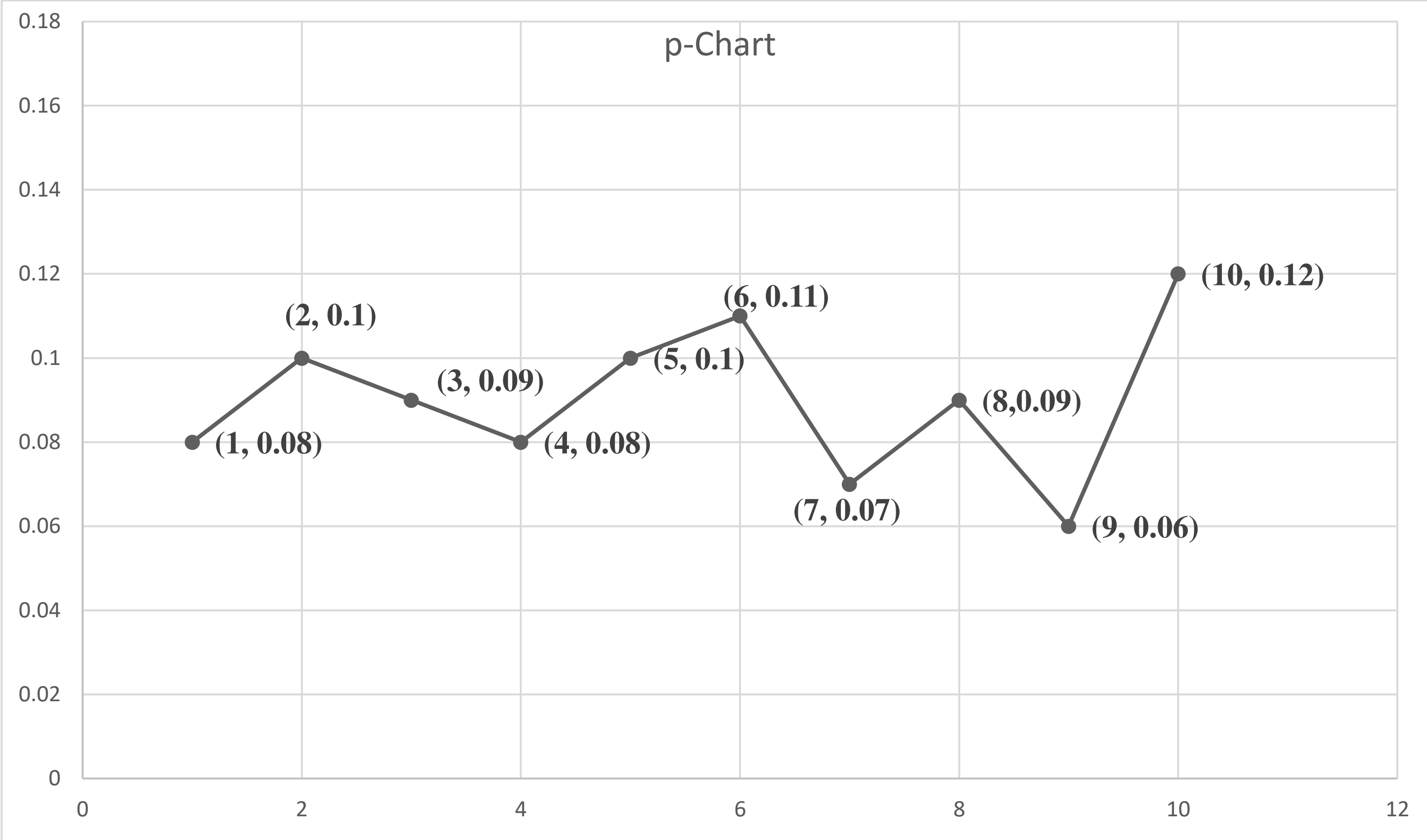
UCL= $\bar{P} + 3\sqrt{\dfrac{\bar{P}(1-\bar{P})}{n}}$

$= 0.09 + 3\sqrt{\dfrac{0.09(0.91)}{100}} = 0.177$

UCL= $\bar{P} - 3\sqrt{\dfrac{\bar{P}(1-\bar{P})}{n}}$

$= 0.09 - 3\sqrt{\dfrac{0.09(0.91)}{100}} = 0.003$

11



Since all the sample points lie within the control lines, the process is under control.

3. The data given below are the number of defectives in 10 samples of 100 items each. Construct a p-chart and an np-chart and comment on the results.

Sample No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
No. of defectives	**6**	**16**	**7**	**3**	**8**	**12**	**7**	**11**	**11**	**4**

Solution:

Sample size is constant for all samples, n=100.

Total no. of defectives = 6 + 16+7+3+8+12+7+11+11+4= 85

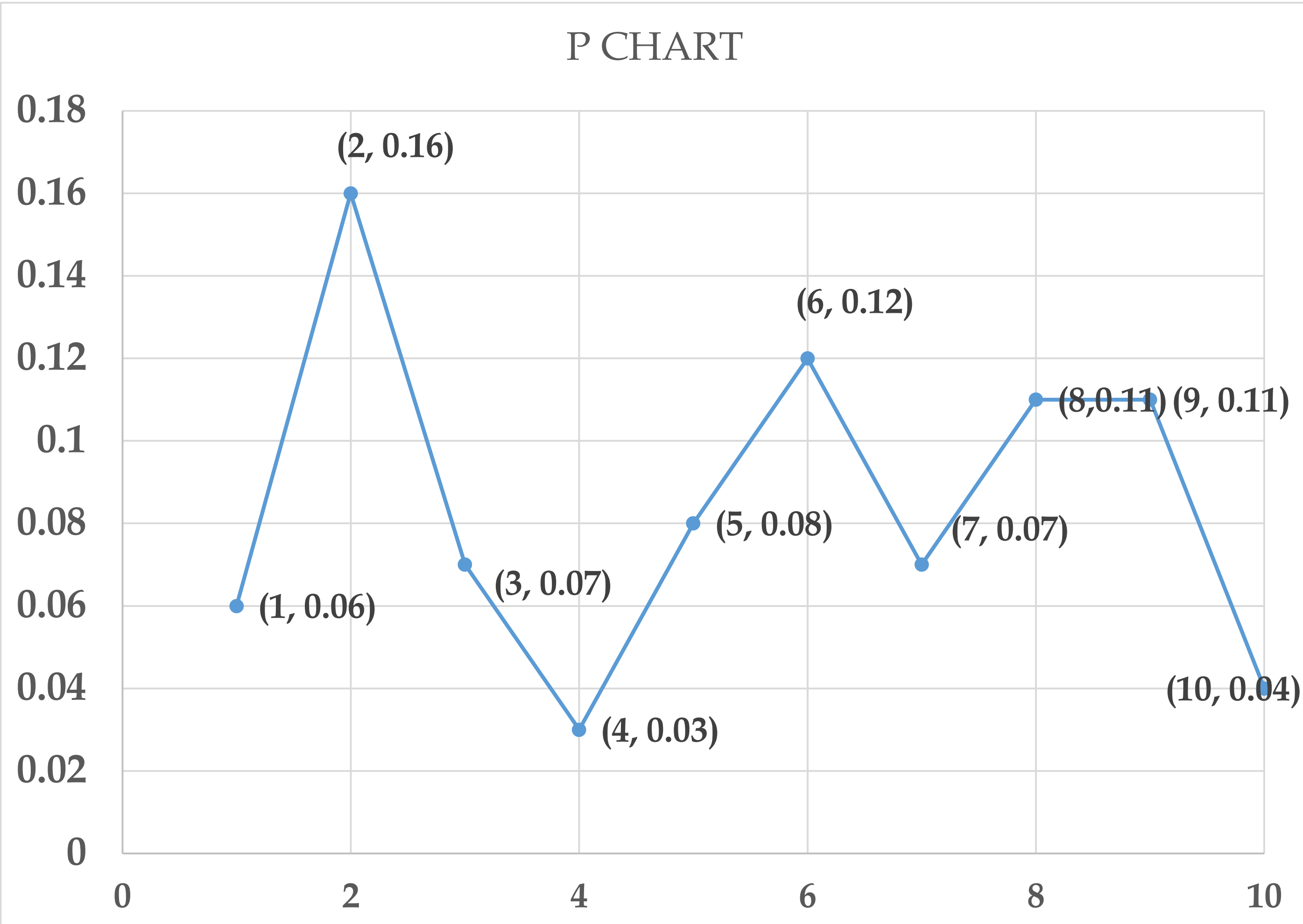
Total no. Inspected= 10 x 100 = 1000

Average fraction defective = $\bar{p} = \frac{\text{Total no. of defectives}}{\text{Total no. of items inspected}} = \frac{85}{1000} = 0.085$

For p-chart:

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.085 - 3\sqrt{\frac{(0.085)(0.915)}{100}} = 0.0013$$

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.085 + \left(\sqrt{\frac{(0.085)(0.915)}{3}}\right) = 0.1687$$



Conclusion:

All these values are less than UCL=0.1687 and greater than LCL=0.0013. In the control chart, all sample points lie within the control limits. Hence, the process is under statistical control.

For np-chart:

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

$$= n\left[\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

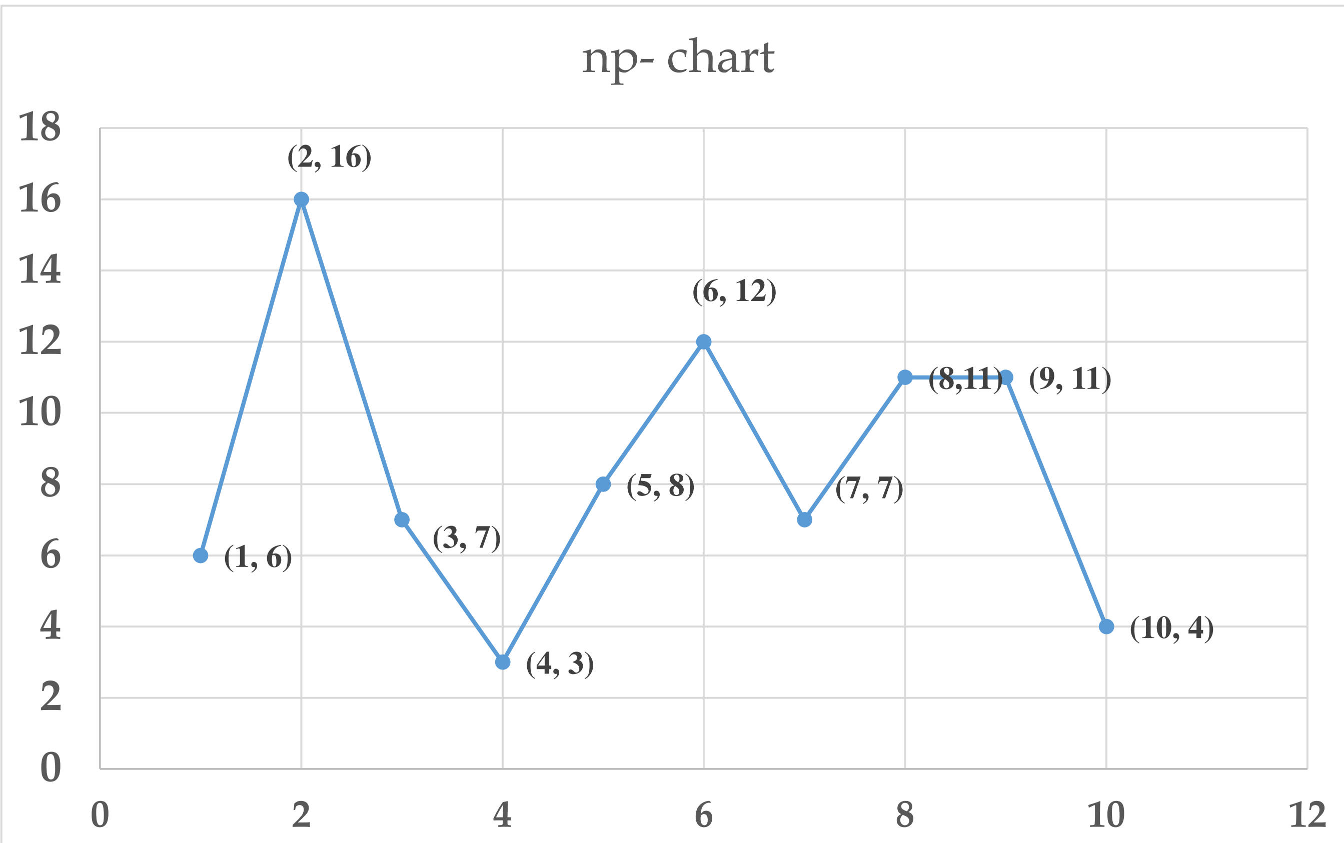
$$= 100(0.1687) = 16.87$$

$$n\bar{p} \quad = 100(0.085) = 8.5$$

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

$$= n\left[\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

$$= 100(0.0013) = 0.13$$



Conclusion:

All the values of number of defectives in the table lie between 16.87 and 0.13. Hence, the process is under control even in np-chart.