Hypothesis Testing :-

Hypothesis :- Hypothesis in statistics is an assumption or claim about a population or a situation that we want to test. It is like an educated guess based on some observation, and we use data to check whether it is true or false.

For example - • A company might hypothesis, "our new advertisement will increase sales"

• A scientist might hypothesis, "This medicine reduces blood pressure ~~more efficiently~~ effectively than the current One".

By collecting and analyzing data, we can test whether these hypothesis are correct.

* Types of Hypothesis :-

1) NULL hypothesis :- (H_0) is the default assumption that nothing new or different is happening. It's like saying "There is no effect or no difference". For example "The new medicine has no effect on blood pressure".

It is a claim or statement about population parameter that is assumed to be true until ~~as~~ it is declared to be false.

• Alternative Hypothesis :— (H_1) is the opposite of the null hypothesis. It's what you are trying to prove, and it suggest that "something is happening, or there's is a difference".

For example,

• "The new medicine lowers blood pressure".

• "The new teaching method is more effective".

		H_0	H_1
T	H_0	...	Type 1 error (seller's error)
F	H_1	Type 2 error (buyer's error)	

Hypothesis Testing

• Test for significant for attributes

• ~~The~~ Test for significant for variables - Large sample / Small sample

Attributes :—

① Test for no. of success

Standard error (SE) of no. of success = $\sqrt{npq}$

n = size of sample
p = probability of success in each trail
q = probability of failure.

② Test for proportion of success = $\sqrt{pq/n}$

Difference - $\pm 3\sqrt{pq/n}$

③ Test for difference b/w proportion

$$S.E(P_1 - P_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \text{ , if } \frac{P_1 - P_2}{SE} < 1.96 SE$$

$$p = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2}$$

SHRI VAISHNAV VIDHYAPEETH VISHWAVIDHYALAYA

Lab Code | Roll No. | Exp. No. | Page

Two tailed test for different b/w two samples

$S.E = \sqrt{a^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ (both samples are from one population)

$SE = \sqrt{\left(\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}\right)}$

Small sample Assumptions:—

1) The Population is normally distributed
2) The value given by the Sample data are sufficiently close to the population value and can be used in their place.

$SE \text{ of median} = \frac{a}{\sqrt{N}}$

$SE \text{ of coefficient of } \cancel{\text{card}} \text{ correlation} = \frac{1-r^2}{\sqrt{N}}$

* t test / t distribution:—

$t = \frac{(\bar{X}-\mu)\sqrt{n}}{S}$

~~μ = sample mean~~
~~$\bar{X}$ = population mean~~

$S = \sqrt{\frac{\Sigma(x-\bar{X})^2}{n-1}}$

$\bar{x}$ = mean of sample

μ = population mean

Properties of t-test distribution

1) t-distribution range from $-\infty$ to ∞
2) Like normal distribution, t distribution is symmetrical with mean, median and mode equal to zero.

Testing of Hypothesis

i) z-test : sample size $(n) > 30$
ii) t-test : sample size $(n) \leq 30$

Alternative Hypothesis

$H_1 = \mu < \mu_0 \Rightarrow$ left tailed test $= \bar{x} < \mu_0$
$H_1 = \mu > \mu_0 \Rightarrow$ Right tailed test $= \bar{x} > \mu_0$
$H_1 = \mu \neq \mu_0 \Rightarrow$ Two ~~tai~~ tailed test

T test when we have more than 1 sample/data

$$s = \sqrt{\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2}{n_1+n_2-2}}$$

$$t = \frac{\text{difference}}{SE} \times \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

degree of freedom $= n_1+n_2-2$

Techniques of identifying one tailed & two tailed

less than, more than, lower, higher, increase, decrease, Smaller than, taller than, greater than, inferior, superior, claims that, at least, at most, Only, improved $\Rightarrow$ one tailed test

Otherwise $\Rightarrow$ Two tailed test

* F-test :-

$$F = \frac{S_1^2}{S_2^2} \qquad S_1 = \frac{\Sigma(x_A - \bar{x}_A)^2}{n-1}$$

$$S_2 = \frac{\Sigma(x_B - \bar{x}_B)^2}{n-1}$$

Level of signiface

$u_1 = n-1$

$u_2 = n-1$

Date: / /
Page:

* Control Chart :- (For variables)

Sample No. ↓	Measurement per Sample X (in cm) 1	2	3
1	.488	.489	.505
2	.494	.495	.499
3	.498	.515	.487
4	.492	.509	.514
5	.490	.508	.499

Calculate control unit for $\bar{x}$, R charts, Draw $\bar{x}$, R charts and examine whether the process is in Statistical control? [$A_2 = 1.02$, $D_4 = 2.57$, $D_3 = 0$, for $n=3$]

n=3

x_1	x_2	x_3	x_4	x_5
.488	.494	.498	.492	.490
.489	.495	.515	.509	.508
.505	.499	.487	.514	.499
$\bar{x}_1 = 0.494$	$\bar{x}_2 = 0.496$	$\bar{x}_3 = 0.5$	$\bar{x}_4 = 0.505$	$\bar{x}_5 = 0.499$

Range = max − min

R_1 = Range (x_1) = .505 − .488 = 0.017
R_2 = Range (x_2) = .499 − .494 = 0.005
R_3 = Range (x_3) = .515 − .487 = 0.028
R_4 = Range (x_4) = .514 − .492 = 0.022
R_5 = Range (x_5) = .508 − .490 = 0.018

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5}{5} = \frac{2.494}{5} = \cancel{0.48}\ \boxed{0.4988}$$

Date:
Page:

$\boxed{A_2 = 1.02}$ $R_1 = 0.017$ $R_4 = 0.022$

$\boxed{D_3 = 0}$ $R_2 = 0.005$ $R_5 = 0.018$

$\boxed{D_4 = 2.57}$ $R_3 = 0.028$

$$\bar{R} = \frac{R_1 + R_2 + R_3 + R_4 + R_5}{5} = \frac{0.090}{5}$$

$$= 0.018$$

$\boxed{\bar{R} = 0.018}$ $\boxed{\bar{X} = 0.4988}$

Trail control limit for $\bar{x}$ - chart

$CL = \bar{\bar{X}}$

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

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

$\boxed{CL = \bar{\bar{X}} = 0.4988}$

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

$= 0.4988 + 1.02 \times 0.018$

$= 0.4988 + 0.01836$

$\boxed{UCL = 0.51716}$

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

$= 0.4988 - 1.02 \times 0.018$

$= 0.4988 - 0.01836$

$\boxed{LCL = 0.4804}$

control limit for R chart

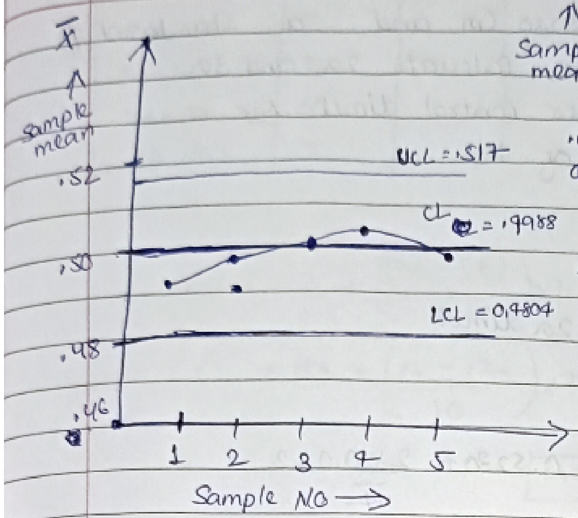
$CL = \bar{R} = 0.018$

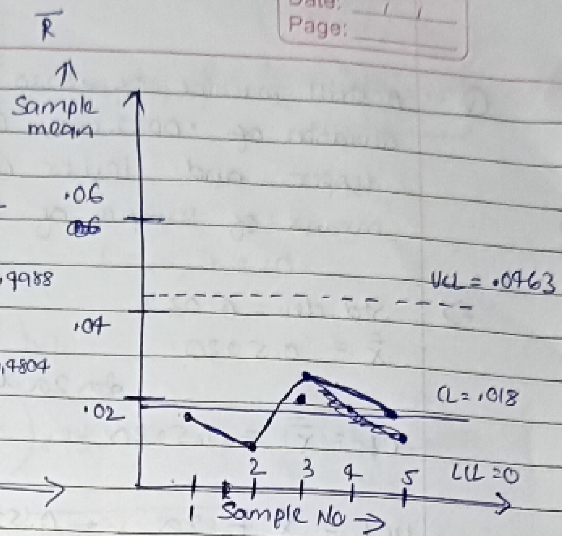
$UCL = D_4 \bar{R} = 2.57 \times 0.018 = 0.0462$

$LCL = D_3 \bar{R} = 0$

$\boxed{UCL = 0.0462}$

$\boxed{LCL = 0}$





Yes, Process is in statistical control

Date: / /
Page:

Q mean deviation

class interval	Frequency (f)	mid (x)	fx	$\lvert x-\bar{x}\rvert$
0 - 10	4	5	20	10.33
10 - 20	6	15	90	0.33
20 - 30	5	25	120	9.67
	$\Sigma f = 15$		$\Sigma fx = 230$	

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{230}{15} = 15.33$$

$$\boxed{\bar{x} = 15.33}$$

$f\lvert x-\bar{x}\rvert$
41.32
1.98
48.35
$\Sigma f\lvert x-\bar{x}\rvert = 91.65$

$$MD = \frac{\Sigma f\lvert x-\bar{x}\rvert}{\Sigma f} = \frac{91.65}{15}$$

$$\boxed{\text{mean deviation} = 6.11}$$

standard deviation :-

• for Grouped Data

$$SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

• For ungrouped Data

$$SD = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$

For ungrouped

Q $\underline{SD}$:- 10, 20, 30, 40, 50

$$\text{mean} = \frac{10+20+30+40+50}{5}$$

$$= \frac{150}{5}$$

$$\boxed{\text{mean} = 30}$$

variance

$$= (x-\bar{x})^2$$

$$= \frac{(-20)^2+(-10)^2+(0)^2+(10)^2+(20)^2}{5}$$

$$= \frac{400+100+100+400}{5}$$

$$= \frac{1000}{5} \quad 200$$

variance = 200

$$\text{Standard dev} = \sqrt{\text{variance}}$$

$$= \sqrt{200}$$

$$\boxed{\text{Std. dev} = 14.142}$$

For Grouped

class Inteval	Freq (f)	mid (x)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
0-10	3	5	15	81	243
10-20	5	15	75	1	5
20-30	2	25	50	121	242
	$\Sigma f = 10$		$\Sigma fx = 140$	$\boxed{\cancel{203}}$	490

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{140}{10}$$

$\boxed{\cancel{\bar{x} = 10}}$ $\boxed{\bar{x} = 14}$

$$\text{variance} = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f}$$

$$= \frac{490}{10}$$

variance = 49

$$\text{std. dev} = \sqrt{49} = 7$$

$\boxed{\text{std. dev} = 7}$

Q co-efficient of variation

class int	freq (f)	mid (m)	fxm	fm^2
0-10	5	5	25	125
10-20	8	15	120	1800
20-30	15	25	375	9375
30-40	16	35	560	19600
40-50	6	45	270	12150
	$\boxed{N = \Sigma f = 50}$		$\boxed{\Sigma fm = 1350}$	$\boxed{\Sigma fm^2 = 43050}$

$$\text{Std. dev} = \sqrt{\frac{fm^2}{N} - \left[\frac{\Sigma fm}{N}\right]^2} = \sqrt{\frac{43050}{50} - \left[\frac{1350}{50}\right]^2}$$

$$= \sqrt{\frac{43050}{50} - (27)^2}$$

$$= \sqrt{861 - 729} = \sqrt{132}$$

$\boxed{\text{Std} = 11.489}$

$$\bar{x} = \frac{1350}{50}\ 27$$

$$\text{coefficient} = \frac{\text{std}}{\text{mean}} \times 100 = \frac{11.489}{27} \times 100 = \frac{1148.9}{27} = \boxed{42.55}$$

Date:
Page:

Q Co-efficient of variation :-

marks	(f) No. of student	(m)	fm	fm^2
0-10	2	5	10	50
10-20	4	15	60	900
20-30	5	25	125	3125
30-40	9	35	315	11025
40-50	10	45	450	20250
50-60	5	55	275	15125
60-70	15	65	975	63375
$\boxed{N = \Sigma f = 50}$			$\boxed{\Sigma fm = 2210}$	$\boxed{\Sigma fm^2 = 113850}$

$$\text{std. dev} = \sqrt{\frac{\Sigma fm^2}{N} - \left[\frac{\Sigma fm}{N}\right]^2}$$

$$= \sqrt{\frac{113850}{50} - \left(\frac{2210}{50}\right)^2}$$

$$= \sqrt{\frac{113850}{50} - (44.2)^2}$$

$$= \sqrt{2277 - 1953.64}$$

$$= \sqrt{323.36}$$

$$\text{std} = 17.98$$

$$\bar{x} = \frac{\Sigma fm}{N} = \frac{2210}{50}$$

$$\boxed{\bar{x} = 44.2}$$

$$\text{Co. var} = \frac{\text{std}}{\bar{x}} \times 100$$

$$= \frac{17.98}{44.2} \times 100$$

$$= \frac{1798}{44.2}$$

$$\text{Co. var} \; \boxed{= 40.68}$$

Date: / /
Page:

Karl pearson / co-efficient of correlation

x	y	x^2	y^2	xy
1	2	1	4	2
2	3	4	9	6
3	5	9	25	15
4	7	16	49	28
5	8	25	64	40
$\Sigma x = 15$	$\Sigma y = 25$	$\Sigma x^2 = 55$	$\Sigma y^2 = \cancel{102}$ 151	$\Sigma xy = 91$

$N = 5$

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{5 \times 91 - 15 \times 25}{\sqrt{5 \times 55 - (15)^2} - \sqrt{5 \times \cancel{102}\,151 - (25)^2}}$$

$$= \frac{455 - 375}{\sqrt{275 - 225}\ \cancel{\sqrt{510 - 625}}\ \sqrt{755 - 625}}$$

$$= \frac{80}{\sqrt{50}\sqrt{130}} = \frac{80}{7.07 \times 11.40} = \frac{80}{80.598}$$

$$\boxed{r = 0.99}$$

Rank correlation (spearman)

Rank ~~corret~~ correlation :-

X	Y	(X) R_1	(Y) R_2	$d = R_1 - R_2$
10	8	1	1	
20	12	2	2	
30	28	3	3	
40	34	4	4	
50	42	5	5	

Age of Husband (H)	Age of wife (W)	(H) R_1	(W) R_2	$d = R_1 - R_2$	d^2
23	21	1	1	0	0
27	23	2	3	-1	1
28	27	3	7	-4	16
29	29	4	9	-5	25
30	32	5	10	-5	25
31	22	6	2	4	16
33	25	7	5	2	4
35	24	8	4	4	16
36	26	9	6	3	9
39	28	10	8	2	4
				15 - 15	
				$\boxed{\Sigma d = 0}$	$\boxed{\Sigma d^2 = 116}$

$$\boxed{\text{Rank correlation} = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}}$$

$$= 1 - \frac{6 \times 116}{10(10^2-1)} = 1 - \frac{6 \times 116}{10 \times 99}$$

$$= 1 - \frac{696}{990} \qquad = 1 - \frac{96}{990}$$

$$= 1 - .070 \qquad = 1 - 0.096$$

$$\boxed{= 0.3} \qquad = 0.904$$

Date: / /
Page:

$\boxed{N=8}$

Q

x	y	xy	x^2	y^2
3	65	195	9	4225
4	70	280	16	4900
6	80	480	36	6400
5	75	375	25	5625
7	85	595	49	7225
8	90	720	64	8100
10	95	950	100	9025
9	88	792	81	7744
$\boxed{\Sigma x = 52}$	$\boxed{\Sigma y = 648}$	$\boxed{\Sigma xy = 4387}$	$\boxed{\Sigma x^2 = 380}$	$\boxed{\Sigma y^2 = 53244}$

$\boxed{\bar{y}=81}$ $\boxed{\bar{x}=6.5}$

y on x

b_{yx}

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

$$b_{yx} = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{8 \times 4387 - 52 \times 648}{8 \times 380 - (52)^2}$$

$$= \frac{35096 - 33696}{3040 - 2704}$$

$$= \frac{1400}{336} = 4.16$$

$\boxed{b_{yx} = 4.16}$

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

$(y-81) = 4.16(x-6.5)$

$y-81 = 4.16x - 27.04$

$y = 4.16x - 27.04 + 81$

$\boxed{y = 4.16x + 53.96}$

x on y

b_{xy}

$(x-\bar{x}) = b_{xy}(y-\bar{y})$

$$b_{xy} = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{8 \times 4387 - 52 \times 648}{8 \times 53244 - (648)^2}$$

$$= \frac{35096 - 33696}{425952 - 419904}$$

$$= \frac{1400}{6048} = 0.231$$

$\boxed{b_{xy} = 0.231}$

$(x-\bar{x}) = b_{xy}(y-\bar{y})$

$(x-6.5) = 0.231(y-81)$

$x - 6.5 = 0.231y - 18.71$

$x = 0.231y - 18.71 + 6.5$

$\boxed{x = 0.231y - 12.21}$

5

Q2. $n = 50$, mean $= 35$,
Std dev $= 5$,

co-efficient of variation

$$= \frac{\text{Std. dev}}{\text{mean}} \times 100$$

$$= \frac{5}{35} \times 100$$

$$= \frac{1}{6} \times 100$$

$$= 16.666$$

10.5
9.5
1.0

Q4.

Physics	Math	R_1 Phy	R_2 M	$d = R_1 - R_2$	
20	50	2	6	-4	16
25	40	3	4.5	-1.5	2.2
35	60	5	7	-2	4
15	30	1	3	-2	4
40	25	6	2	4	16
50	20	7	1	6	36
35	40	5	4.5	0.5	0.25
60	65	8	8	0	0

$\boxed{\Sigma d = 1}$ $\boxed{\Sigma d^2 = 7}$

$\frac{3+7}{2} = \frac{10}{2}$ $\boxed{5}$

$\boxed{\begin{matrix} m_1 = 2 \\ m_2 = 2 \end{matrix}}$

$3 + 4.5$

$\frac{2+7}{2} = \frac{9}{2}$

$8-2=6$

Date: / /
Page:

$$r = 1 - 6\left[\frac{\Sigma d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12}}{n(n^2-1)}\right]$$

$$= 1-6\left[\frac{78.5 + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12}}{2(2^2-1)}\right]$$

$$= \cancel{10}^{-}1 - 6\left[\frac{78.5 + \frac{2(4-1)}{12} + \frac{2(4-1)}{12}}{2(4-1)}\right]$$

$$= 1-6\left[\frac{78.5 + \frac{\cancel{2}(3)}{\cancel{12}\cancel{6}\,2} + \frac{\cancel{2}(3)}{\cancel{12}\,\cancel{6}\,2}}{2(3)}\right]$$

$$= 1 - 6\left[\frac{78.5 + \frac{1}{2} + \frac{1}{2}}{6}\right] = \cancel{0}^{-}1 - 6\left[\frac{78.5 + \frac{\cancel{2}}{\cancel{2}}1}{6}\right]$$

$$= 1 - 6\left[\frac{79.5}{6}\right]$$

$$= 1 - \frac{13.25}{6}$$

$$= 1 - 2.208$$

$$\boxed{R = -1.208}$$

Chi square test

Q

	(O_{11}) smoker	(O_{12}) Non-smoker	
M	30	20	50
F	15	25	40
	45	45	$\boxed{90}$ = Grand Total

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

For (male, smoker)

$$E_{11} = \frac{\text{Row Total for male} \times \text{column Total for smoker}}{\text{Grand Total}}$$

$$= \frac{50 \times 45}{90} = \frac{2250}{90} = 25$$

For (male, non smoker)

$$= \frac{50 \times 45}{90} = \frac{2250}{90} = 25$$

for (female, smoker)

$$= \frac{40 \times 45}{90} = \frac{\cancel{1800}^{20}}{\cancel{90}} = 20$$

For (female, non smoker)

$$= \frac{40 \times 45}{90} = 20$$

Gender	smoker	Non smoker	
male	25	25	50
female	20	20	40
	45	45	$\boxed{90}$

G	S	N-S
M	25	25
F	20	20

G	S	N-S	
m	25	25	50
F	20	20	40
	45	45	90

calculate chi-square test $(\chi)^2$

$$= \frac{(30-25)^2}{25} = \frac{(5)^2}{25} = \frac{25}{25} = 1$$

$$= \frac{(20-25)^2}{25} = \frac{(-5)^2}{25} = \frac{25}{25} = 1$$

$$= \frac{(15-20)^2}{20} = \frac{(-5)^2}{20} = \frac{25}{20} = 1.25$$

$$= \frac{(25-20)^2}{20} = \frac{(5)^2}{20} = \frac{25}{20} = 1.25$$

$$\chi^2 = 1+1+1.25+1.25$$

$$\boxed{\chi^2 = 4.5}$$

Degree of freedom

$(n-1)(c-1) = (2-1)\times(2-1)$

$= 1\times1$

$= 1$

$\boxed{\alpha = 0.05 \quad df = 1}$

Ftab		Fcal
3.8	<	4.5

~~Null~~ True

One way Anova

	A	B	C	D
Method 1 :	85,	88,	90,	84
method 2 :	78,	82,	80,	85
method 3 :	92,	88,	95,	89
	255	258	265	258

$\bar{x}_A = 85 \quad \bar{x}_B = 86 \quad \bar{x}_C = 88.3 \quad \bar{x}_D = 86$

$$\text{Grand mean} = \frac{85+86+88.3+86}{4}$$

$$\bar{\bar{x}} = \frac{345.3}{4}$$

$$\boxed{\bar{\bar{x}} = 86.3}$$

- SSC

$\bar{x}_A - \bar{\bar{x}}$	$(\bar{x}_A - \bar{\bar{x}})^2$	$(\bar{x}_B - \bar{\bar{x}})$	$(\bar{x}_B - \bar{\bar{x}})^2$	$(\bar{x}_C - \bar{\bar{x}})$	$(\bar{x}_C - \bar{\bar{x}}$
85 − 86.3 = (−1.3)	16.9	86 − 86.3 = −0.3	$0.\cancel{09}$	88.3 − 86.3 = 2	4
85 − 86.3 = (−1.3)	16.9	86 − 86.3 = −0.3	$0.\cancel{09}$	88.3 − 86.3 = 2	4
85 − 86.3 = (−1.3)	16.9	86 − 86.3 = −0.3	$0.\cancel{09}$	88.3 − 86.3 = 2	4

~~SSC = 16.9 + 0 + 4~~

~~SSC = 20.9~~

$\bar{x}_D - \bar{\bar{x}}$	$(x_D - \bar{\bar{x}})^2$
86 − 86.3 = −0.3	0.09
86 − 86.3 = −0.3	0.09
86 − 86.3 = −0.3	0.09

$SSC = 16.9 + 0 + 4 + 0.09$

$$\boxed{SSC = 20.99}$$

SSE :—

$A-\bar{x}_A$	$(A-\bar{x}_A)^2$	$(B-\bar{x}_B)$	$(B-\bar{x}_B)^2$	$(C-\bar{x}_C)$	$(C-\bar{x}_C)^2$
85−85=0	0	88−86=2	4	90−88.3=1.7	=2.89
78−85=−7	49	82−86=−4	16	80−88.3=−8.3	=68.89
92−85=7	49	88−86=2	4	95−88.3=6.7	=44.89
	98		24		116.67

$SSE = 98 + 24 + 116.67 + 14$

$\boxed{SSE = 252.67}$

$\boxed{SSC = 20.99}$

- Ration of f

$F = \frac{MSC}{MSE}$

$= \frac{6.99}{31.58}$

$\boxed{F = 0.22}$

$MSC = \frac{SSC}{C-1} = \frac{20.99}{4-1}$

$= \frac{20.99}{3} =$

$\boxed{MSC = 6.99}$

$MSE = \frac{SSE}{n-C} = \frac{252.67}{12-4}$

$= \cancel{252}$

$= \frac{252.67}{8}$

$\boxed{MSE = 31.258} = 31.258$

Level of significance

$v_1 = C-1 = 4-1 = 3$

$v_2 = n-C = 12-4 = 8$

5% (3,8)

$f_{tab} = 3.86$

Ftab		Fcal
3.86	>	0.22