12/11/18 Mean:

Mean for individual series $\bar{x}$:

$$\bar{x} = \frac{\Sigma x}{N}$$

Mean for discrete case:

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{\Sigma fx}{N \to (\Sigma f)}$$

N

Mean for continuous case:

$$\bar{x} = \frac{\Sigma fm}{\Sigma f}$$

Problem:

1. The following table gives the monthly income of 10 employees in an office.

1780	1760	1690	1750	1840	1920	1100	1810
					1050	1950.	

Individual case:

Mean $\bar{x} = \frac{\Sigma x}{N}$

$$= \frac{1780+1760+1690+1750+1840+1920+1100+1810+1050+1950}{10}$$

$$= \frac{16650}{10}$$

$$\boxed{\text{Mean } \bar{x} = 1665}$$

2. From the following data of marks obtained by 60 students of class. calculate the arithematic mean.

Marks(x)	20	30	40	50	60	70
No of Students (f)	8	12	20	10	6	4

discrete case:

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

$$= \frac{2460}{60}$$

$$\boxed{\bar{x} = 41}$$

x	f	fx
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	60	2460

3. From the following data complete the arithematic mean.

Marks (x)	0-10	10-20	20-30	30-40	40-50	50-60
No of Students (f)	5	10	25	30	20	10

continuous case:

$$\bar{x} = \frac{\Sigma fm}{\Sigma f}$$

$$= \frac{3300}{100}$$

$$\boxed{\bar{x} = 33}$$

x	f	m	fm
0-10	5	5	25
10-20	10	15	150
20-30	25	25	625
30-40	30	35	1050
40-50	20	45	900
50-60	10	55	550
	100		3300

Combined arithmatic mean :

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1+N_2}$$

$$\bar{X}_{123} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{X}_3}{N_1+N_2+N_3}$$

N_1 - no of item in the 1st group

N_2 - No of itemea in the 2nd group

N_3 - No of item in the 3rd group

$\bar{x}_1$ - mean of 1st group

$\bar{x}_2$ - mean of 2nd group

$\bar{x}_3$ - mean of 3rd group.

Problem :

1. The mean height of 25 male workers in a factory is 61 cm and the mean height of 35 female workers in the same factory is 58 cm. Find the combined mean height of 60 workers in the factory.

$N_1 = 25$ $\quad \bar{X}_1 = 61$ cm

$N_2 = 35$ $\quad \bar{X}_2 = 58$ cm.

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1+N_2}$$

$$= \frac{1525 + 2030}{60}$$

$\boxed{\bar{X}_{12} = 59.25\text{cm}}$

Median :

Median for individual series :

$$\text{Median} = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item.}$$

Median for discrete case :

$$\text{Median} = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item.}$$

Here $N = \Sigma f$

Median for continuous case :

$$\text{Median} = \text{size of } \left(\frac{N}{2}\right)^{th} \text{ item.}$$

$$\text{Median} = L + \left[\frac{\frac{N}{2} - c.f}{f}\right] \times i.$$

Here L is the lower limit of median class

cf is the cummulative frequency of the class preceding the medium class

f is the frequency of the class

i is the interval.

N - Total no of frequency.

1. From the following data of wages of 7 workers. compute the median wages

wages in Rs:

1100	1150	1080	1120
1200	1160	1400	

Ascending order:

wages: 1080 1100 1120 1150 1160 1200 1400.

individual case:

Median = size of $\left(\frac{N+1}{2}\right)^{th}$ item.

= size of $\left(\frac{7+1}{2}\right)^{th}$ item.

= size of $(4)^{th}$ item.

$\boxed{\text{Median} = 1150}$

2. Obtain the value of median from the following data.

391	384	591	407	672	522	~~777~~	753	2488	1490

Ascending order

384 391 407 522 591 672 753 777

1490 2488

individual case:

$$\text{Median} = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}$$

$$= \text{size of } \left(\frac{10+1}{2}\right)^{th} \text{ item.}$$

$$= \text{size of } (5.5)^{th} \text{ item.}$$

$$= \frac{591+672}{2}$$

$$\boxed{\text{Median} = 1263/2}$$

$$\boxed{\text{Median} = 631.5}$$

---

3. From the following data find the value of median

Incom	1000	1500	800	2000	2500	1800
No of persons	24	26	16	20	6	30

Income : N cf

800	16	16
1000	24	40
1500	26	66 → choose 2nd value
1800	30	96
2000	20	116
2500	6	122
	N= 122	

individual case:

Median = size of $\left(\frac{N+1}{2}\right)^{th}$ item

= size of $\left(\frac{10+1}{2}\right)^{th}$ item.

= size of $(5.5)^{th}$ item.

= $\frac{591+672}{2}$

$\boxed{\text{Median} = 1263/2}$

$\boxed{\text{Median} = 631.5}$

3. From the following data find the value of median

Income	1000	1500	800	2000	2500	1800
No of persons	24	26	16	20	6	30

Income : N Cf

800	16	16
1000	24	40
1500	26	66 → Choose 2nd value
1800	30	96
2000	20	116
2500	6	122
	N = 122	

discrete case:

$$\text{Median} = \text{size of } \left(\frac{\overset{N}{\sum f}+1}{2}\right)^{th} \text{item}$$

$$= \text{size of } \left(\frac{123\ \cancel{+1}}{2}\right)^{th} \text{item}$$

$$= \text{size of } (61.5)^{th} \text{item.}$$

$$\text{Median} = 1500$$

Q. Calculate the median from the following data.

marks	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No of Students	7	15	24	31	42	30	26	15	10

x	f	cf
5-10	7	7
10-15	15	22
15-20	24	46
20-25	31	77 → cumulative freq: cf
lower limit 25-30	42 corresponding freq f	119
30-35	30	149
35-40	26	175
40-45	15	190
45-50	10	200 = N

continuous case : size of $\left(\frac{N}{2}\right)^{th}$ item.

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$\text{Median} = \left(\frac{200}{2}\right)^{th} th.$$

$$= 100^{th} \text{ item.}$$

$$\text{Median} = 25 + \left(\frac{100 - 77}{42}\right) \times 5$$

$$= 25 + \frac{23 \times 5}{42}$$

$$= 25 + \frac{115}{42}$$

$$\boxed{\text{Median} = 27.73}$$

5. Calculate the median from the following data

x :	110-120	120-130	130-140	140-150	150-160	160-170	170-180	180-190	190-200
f :	6	25	48	72	116	60	38	22	3

x	f	c.f.
110-120	6	6
120-130	25	31
130-140	48	79
140-150	72	151 cf
150-160 L	116 f	267
160-170	60	327
170-180	38	365
180-190	22	387
190-200	3	390

Continuous case:

$$\text{Median} = \text{size of } \left(\frac{N}{2}\right)^{th} \text{ item}$$

$$= \text{size of } \left(\frac{390}{2}\right)^{th} \text{ item}$$

$$= \text{size of } (195)^{th} \text{ item.}$$

$$\text{Median} = L + \left(\frac{\frac{N}{2} - Cf}{f}\right) \times i$$

$$= 150 + \left(\frac{195 - 151}{116}\right) \times 10.$$

$$= 150 + \frac{440}{116}$$

$$\boxed{\text{Median} = 153.79.}$$

---

6. An incomplete distribution is given below:

Variable	0-10	10-20	20-30	30-40	40-50	50-60	60-70
frequency	10	20	f_1	40	f_2	25	15

Given that the median values 35 and total frequencies 170. Find the missing frequency

~~Contd~~

x	f	c.f.
0-10	10	10
10-20	20	30
20-30	f_1	$30+f_1$
30-40	40	$70+f_1$
40-50	f_2	$70+f_1+f_2$
50-60	25	$95+f_1+f_2$
60-70	15	$110+f_1+f_2 = N$

Median = 35

Total frequency = 170.
(N)

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$= 30 + \left(\frac{85 - 30 - f_1}{4\cancel{0}}\right) \times 1\cancel{0}$$

$$35 = 30 + \left(\frac{55 - f_1}{4}\right)$$

$$35 = \frac{120 + 55 - f_1}{4}$$

$$140 - 120 - 55 = -f_1$$

$$\boxed{f_1 = 35}$$

$$110 + f_1 + f_2 = 170$$

$$110 + 35 + f_2 = 170$$

$$\boxed{f_2 = 25}$$

Missing frequency

$$\boxed{f_1 = 35}$$

$$\boxed{f_2 = 25}$$

7. Calculate the median from the following data

Weight (in gms)	410-419	420-429	430-439	440-449	450-459	460-469	470-479
No of apples	14	20	42	54	45	18	7

Ans: 443.94

x	f	c.f.	X
410-419	14	14	409.5 - 419.5
420-429	20	34	419.5 - 429.5
430-439	42	76 c.f.	429.5 - 439.5
440-449	54 f	130	439.5 - 449.5
450-459	45	175	449.5 - 459.5
460-469	18	193	459.5 - 469.5
470-479	7	200 = N	469.5 - 479.5

Continuous case :

Median = size of $\left(\frac{N}{2}\right)^{th}$ item.

= size of $\left(\frac{200}{2}\right)^{th}$ item.

= size of $(100)^{th}$ item.

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$= 439.5 + \left(\frac{100 - 76}{54}\right) \times 10.$$

$$= 439.5 + \frac{240}{54}$$

$$\boxed{\text{Median} = 443.944.}$$

Mode continuous case :

$$\text{Mode} = L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$$

L - lower limit.

f_1 - frequency

f_0 - frequency of preceeding model

f_2 - frequency of succeeding the class

i - class interval.

Mode:

A mode is a value of in series of observations of items. It occurs with greatest frequency.

1. Calculate the mode from the data

10	27	24	12	27	27	20	18	15	30

Mode = 27 → repeated no's.

2. Calculate the mode:

x :	1	2	3	$\boxed{4}$	5	6	7	8
f :	4	9	16	$\boxed{25}$	22	15	7	3

Mode : 4.

① Find $L^{-1}\left[\frac{s}{(s^2+a^2)^2}\right]$, using convolution theorem.

② Find the $L^{-1}\left[\frac{s^2}{(s^2+a^2)(s^2+b^2)}\right]$, using Partial fraction method.

③ Find the Laplace transform of

$$f(t)=\begin{cases} E, & 0<t<T/2 \\ -E, & T/2<t<T \end{cases} \text{ and } f(T+t)=f(t)$$

④ Find the L.T of

$$f(t)=\begin{cases} \sin\omega t ; & 0<t<\pi/\omega \\ 0 ; & \pi/\omega<t<\frac{2\pi}{\omega} \end{cases}$$

⑤ Find the L.T of

$$f(t)=\begin{cases} t ; & 0<t<b \\ 2b-t ; & b<t<2b. \end{cases}$$

⑥ Prove that $e^{-\frac{x^2}{2}}$ is self reciprocal under FCT.

Mode in continuous case:

$$\text{Mode} = L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$$

L - lower limit.

f_1 - frequency

f_0 - frequency of preceeding model class

f_2 - frequency of succeeding model class

i - class interval.

Find mean, median, mode.

x	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45
f	7	10	16	30	24	17	10	5	1

x	f	x	m	fm	cf
1-5	7	0.5 - 5.5	3	21	7
6-10	10	5.5 - 10.5	8	80	17
11-15	16 f_0	10.5 - 15.5	13	208	33 cf
16-20	30 f_1	15.5 L - 20.0	18	540	63
21-25	24 f_2	20.0 - 25.5	23	552	87
26-30	17	25.5 - 30.5	28	476	104
31-35	10	30.5 - 35.5	33	330	114
36-40	5	35.5 - 40.5	38	190	119
41-45	1	40.5 - 45.5	43	43	120
	120			2440	

$$\text{Mean} = \frac{\Sigma fm}{\Sigma f}$$

$$\bar{x} = \frac{2440}{120}$$

$$= 20.33$$

Median = size of $\left(\frac{N}{2}\right)^{th}$ item.

$= \text{size of } \left(\frac{120}{2}\right)^{th}$ item.

= size of 60^{th} item.

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$= 15.5 + \left(\frac{60-33}{30}\right) \times 5$$

$$= 15.5 + \frac{27}{30} \times 5$$

$$= 15.5 + 4.5$$

$$= 20.$$

Mode $= L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times i$

$$= 15.5 + \left[\frac{30 - 16}{2(30) - 16 - 24}\right] \times 5$$

$$= 15.5 + \left[\frac{14}{20}\right] \times 5$$

Calculate the mode for the following data.

Marks	10	15	20	25	30	35	40
No of student	8	12	36	25	28	18	9.

X	f I	II	III	IV	V	VI
10	8	20				
15	12		48	56		
20	36	61			73	
25	25		53			89
30	28	46		71		
35	18		27		55	
40	9					

(i)

(ii) 1 − 1 → 1

(iii) repeated no.

	10	15	20	25	30	35	40
I			1	1			
II			1	1			
III				1	1		
IV		1	1	1	1	1	
V		1	1	1			
VI			1	1	1		
		1	4	5 ↑	3	1	

MODE = ~~5~~ 25

Find the mode of the following question.

weight	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180
no of persons	4	6	20	32	33	17	8	2

Grouping table:

x	f_I	II	III	IV	V	VI
100-110	4	10				
110-120	6		26	30		
120-130	20	$\boxed{52}$			$\boxed{58}$	
130-140	32		$\boxed{65}$			$\boxed{85}$
140-150	33	50		$\boxed{82}$		
150-160	17		25		$\boxed{58}$	
160-170	8	10				27
170-180	2					

	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180
I					1			
II			1	1				
III				1	1			
IV				1	1	1		
V		1	1	1	1	1	1	
VI			1	1	1			
		1	3	5↑	5↑	2	1	Bimodal method.

$\boxed{\text{Bimodal Mode} = 3\,\text{Median} - 2\,\text{Mean}}$

x	f	c.f	m	fm
100-110	4	4	105	420
110-120	6	10	115	690
120-130	20	30	125	2500
130-140	32	62	135	4320
140-150	33	95	145	4785
150-160	17	112	155	2635
160-170	8	120	165	1320
170-180	2	122	175	350
	Σ 122		Σ 1120	Σ 17020

Median = size of $\left(\frac{N}{2}\right)^{th}$ item.

$= \left(\frac{122}{2}\right)^{th}$ item.

$= 61^{th}$ item.

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$= 130 + \left(\frac{61 - 30}{32}\right) \times 10$$

$$= 130 + \frac{31}{32} \times 10$$

$$= 130 + 9.6875$$

$$\boxed{\text{Median} = 139.6875.}$$

$$\text{Mean} = \frac{\Sigma fm}{\Sigma f}$$

$$= \frac{17020}{122}$$

$$\boxed{\text{Mean} = 139.508}$$

$$\text{Mode} = 3\text{Median} - 2\text{Mean}$$

$$= 3(139.6875) - 2(139.508)$$

$$= 419.0625 - 279.016$$

$$\boxed{\text{Mode} = 140.0465}$$

$$\text{Median} = \text{size of } \left(\frac{N}{2}\right)^{th} \text{ item.}$$

$$= \left(\frac{122}{2}\right)^{th} \text{ item.}$$

$$= 61^{th} \text{ item.}$$

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$= 130 + \left(\frac{61 - 30}{32}\right) \times 10$$

$$= 130 + \frac{31}{32} \times 10$$

$$= 130 + 9.6875$$

$$\boxed{\text{Median} = 139.6875.}$$

$$\text{Mean} = \frac{\Sigma fm}{\Sigma f}$$

$$= \frac{17020}{122}$$

$$\boxed{\text{Mean} = 139.508}$$

$$\text{Mode} = 3\text{Median} - 2\text{Mean}$$

$$= 3(139.6875) - 2(139.508)$$

$$= 419.0625 - 279.016$$

$$\boxed{\text{Mode} = 140.0465}$$

The median and mode of following distribution are Rs. 33.50 and Rs. 34. Find values f_3, f_4, f_5.

Also total frequency is 230.

Wages (in Rs)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	4	16	f_3	f_4	f_5	6	4

x	f	cf
0-10	4	4
10-20	16	20
20-30	f_3 f_0	$20+f_3$ (c.f)
30-40	f_4 f_1	$20+f_3+f_4$
40-50	f_5 f_2	$20+f_3+f_4+f_5$
50-60	6	$26+f_3+f_4+f_5$
60-70	4	$30+f_3+f_4+f_5$

$$Median = L + \left(\frac{\frac{N}{2} - C.f}{f}\right) \times i$$

$$= 30 + \left(\frac{115 - (20 + f_3)}{f_4}\right) \times 10$$

$$33.50 = 30 + \left(\frac{95 - f_3}{f_4}\right) \times 10$$

$$3.50 = \frac{(95 - f_3) \times 10}{f_4}$$

$$3.5f_4 = 950 - 10f_3$$

$$\boxed{f_4 = 271.4 - 2.85f_3} \text{ — (1)}$$

$$Mode = L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$$

$$\boxed{Mode = 34}$$

$$34 = 30 + \left[\frac{f_4 - f_3}{2f_4 - f_3 - f_5}\right] \times 10$$

$$4 = \left[\frac{f_4 - f_3}{2f_4 - f_3 - f_5}\right] \times 10$$

$$4 = \left[\frac{271.4 - 2.85f_3 - f_3}{2(271.4 - 2.85f_3) - (f_3 + f_5)}\right] \times 10 \text{ — (2)}$$

$$N = 230 = 30 + f_3 + f_4 + f_5$$

$$f_3 + f_5 = 230 + 30 - f_4$$

$$\boxed{f_3 + f_5 = 200 - f_4} \text{ — (3)}$$

substitude (3) in (2),

$$4 = \left[\frac{271.4 - 3.85 f_3}{2(271.41 - 2.85 f_3 - 200 + f_4}\right] \times 10$$

$$4[2(271.4) - 2.85 f_3 - 250 + f_4] =$$

$$[271.4 - 3.85 f_3) \times 10.$$

$$4[542.8 - 5.7 f_3 - 200 + 271.4 - 2.85 f_3]$$

$$= 271.4 - 38.5 f_3$$

$$4.3 f_3 = 256.68$$

$$\boxed{f_3 = 59.693}$$

$$f_3 \simeq 60$$

substitute f_3 in (1),

$$f_4 = 271.4 - 2.85 (59.693)$$

$$\boxed{f_4 = 100.4}$$

$$230 = 30 + 60 + 100 + f_5$$

$$\boxed{f_5 = 40}$$

∴ The frequencies are

$$\boxed{f_3 = 59.693}$$

$$\boxed{f_4 = 100.4}$$

$$\boxed{f_5 = 40.}$$

RANGE

$$\text{Range} = L - S$$

$L \rightarrow$ large value

$S \rightarrow$ smallest value

$$\text{coefficient of Range} = \frac{L-S}{L+S}.$$

1. Calculate range and co-efficient from the data.

Items	200	210	208	160	220	250

Largest value $L = 250$

smallest value $S = 160$

$$\text{Range} = L - S$$
$$= 250 - 160$$
$$= 90$$

$$\text{co-efficient of Range} = \frac{L-S}{L+S}$$
$$= \frac{250-160}{250+160}$$
$$= \frac{90}{410}$$
$$= 0.2195$$

2. Calculate the range and co-efficient from the data.

Marks	10-20	20-30	30-40	40-50	50-60
No of students	8	10	12	8	4

Largest value $L = 60$.

smallest value $S = 10$

Range $= L - S$

$= 60 - 10$

$= 50$

co-efficient of range $= \frac{L-S}{L+S}$

$= \frac{60-10}{60+10}$

$= \frac{50}{70}$

$= 0.714$.

MEAN DEVIDIATION.

1. INDIVIDUAL CASE :

Mean deviation $= \frac{\Sigma|D|}{N}$

co-efficient of mean deviation $= \frac{\text{Mean deviation}}{\text{Median}}$

$D = X - A$

$A =$ Median

$N =$ Total no of items.

2. Calculate the range and co-efficient from the data.

Marks	10-20	20-30	30-40	40-50	50-60
No of students	8	10	12	8	4

Largest value $L = 60$.

Smallest value $S = 10$

$$\text{Range} = L - S$$
$$= 60 - 10$$
$$= 50$$

$$\text{co-efficient of range} = \frac{L-S}{L+S}$$
$$= \frac{60-10}{60+10}$$
$$= \frac{50}{70}$$
$$= 0.714.$$

MEAN DEVIDIATION.

1. INDIVIDUAL CASE :

$$\text{Mean deviation} = \frac{\Sigma |D|}{N}$$

$$\text{coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Median}}$$

$D = X - A$

$A =$ Median

$N =$ Total no of items.

Two Dimensional random variable and random process.

Joint probability distribution.

If X and Y are 2 random variable, the probability distribution for their simultaneous occurrences can be represented by a function, $f(x,y)$ (or) $P(X=x, Y=y)$ for any pair of values (x, y) within range of the random variables X and Y. This function is known as Joint probability distribution of X and Y.

$$f(x,y) = P(X=x, Y=y).$$

Two dimensional distribution

(X, Y)

Discrete	Continuous.
Joint (PMF) $P(X=x_i, Y=y_i) = P(x_i, y_j) = P_{ij}$ 1) $P(x=x_i, Y=y_j) \geq 0$ 2) $\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) = 1.$	Joint (PDF). 1) $\iint_R (f(x,y))\,dx\,dy = 1.$ 2) $f(x,y) \geq 0.$
Marginal (PMF) $P_X(x) = \sum_y P(x,y).$ $P_Y(y) = \sum_x P(x,y).$	Marginal (PDF) $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy.$ $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$

Conditional Probability distribution	Conditional Probability distribution
1] $P(X=x_i / Y=y_j) = \frac{P(X \cap Y)}{P(Y)}$. C.P.d [X given Y]. 2] $P(Y=y_j / X=x_i) = \frac{P(X \cap Y)}{P(X)}$ C.P.d [Y given X]	1] $f(x/y) = \frac{f(x,y)}{f_Y(y)}$ 2] $f(y/x) = \frac{f(x,y)}{f_X(x)}$
Independent: $P(X=x_i \cap Y=y_j) = P[X=x_i]\,P[Y=y_j]$	Independent $f(x,y) = f_X(x) \cdot f_Y(y)$.

1] For the following bivariate probability distribution of X and Y, find.

(i) $P(X \le 2, Y=2)$ (ii) $F_X(2)$ (iii) $P(Y=3)$.

(iv) $P(X<3, Y \le 4)$ (v) $F_Y(3)$.

X \ Y	1	2	3	4	$P(X=x)$
1	0.1	0	0.2	0.1	0.4
2	0.05	0.12	0.08	0.01	0.26
3	0.1	0.05	0.1	0.09	0.34.
$P(Y=y)$	0.25	0.17.	0.38	0.2	

(i) $P(X \le 2, Y=2)$:

$$P(x \le 2, Y=2) = P(X=1, Y=2) + P(X=2, Y=2).$$

$$= 0 + 0.12.$$

$$P(X \le 2, Y=2) = 0.12$$

(ii) $F_X(2)$:

$$F_X(2) = P(X \leq 2)$$

$$= P(X=1) + P(X=2)$$

$$= 0.4 + 0.26$$

$$= 0.66$$

(iii) $P(Y=3)$

$$P(Y=3) = 0.2 + 0.08 + 0.1$$

$$= 0.38$$

(iv) $P(X < 3, Y \leq 4)$:

$$= P(X=1, Y \leq 4) + P(X=2, Y \leq 4)$$

$$= 0.4 + 0.26$$

$$= 0.66$$

(v) $F_Y(3)$:-

$$F_Y(3) = P(Y \leq 3)$$

$$= P(Y=1) + P(Y=2) + P(Y=3)$$

$$= 0.25 + 0.17 + 0.38$$

$$F_Y(3) = 0.8$$

2] The joint distribution of X and Y is given by $f(x,y) = \frac{x+y}{21}$, $x = 1,2,3$ and $y = 1,2$. Find the marginal distribution of X and Y.

Solution:

It is discrete with finite value.

X \ Y	1	2	P(X=x)
1	2/21	3/21	5/21
2	3/21	4/21	7/21
3	4/21	5/21	9/21
P(Y=y)	9/21	12/21	1

$f(1,1) = \frac{1+1}{21} \Rightarrow \frac{2}{21} \rightarrow 0.095$

$f(1,2) = \frac{1+2}{2} = \frac{3}{21} = 0.14$

So on

$f($

Marginal distribution of X:

$P(X=1, Y=y_j) = P(1,1) + P(1,2)$

$= 2/21 + 3/21 = 5/21$

$P(X=2, Y=y_j) = P(2,1) + P(2,2)$

$= 3/21 + 4/21 = 7/21$

$P(X=3, Y=y_j) = P(3,1) + P(3,2)$

$= 4/21 + 5/21 = 9/21$

Marginal distribution of Y.

$P(X=x_i, Y=1) = P(1,1) + P(2,1) + P(3,1)$

$= 2/21 + 3/21 + 4/21$

$= 9/21$

$P(X=x_i, Y=2) = P(1,2) + P(2,2) + P(3,2)$

$= 3/21 + 4/21 + 5/21$

$= 12/21$

3] The joint probability mass function of (x,y) is $P(x,y) = K(2x+3y)$, $x = 0,1,2$; $y = 1,2,3$. Find all marginal and conditional probability distribution.

solution

x \ y	1	2	3	P(X=x)
0	3k	6k	9k.	18k
1	5k	8k	11k.	24k
2	7k	10k	13k.	30k
P(Y=y)	15k	24k	33k.	

$P(0,1) = k(2(0)+3(1))$.

$= 3k$

To find the value of k.

$$\sum_{i=0}^{2}\sum_{j=1}^{3} P(x_i, y_j) = 1.$$

$$72k = 1.$$

$$\boxed{k = 1/72.}$$

x \ y	1	2	3	P(X=x)
0	3/72	6/72	9/72	18/72
1	5/72	8/72	11/72	24/72
2	7/72	10/72	13/72	30/72
P(Y=y)	15/72	24/72	33/72	1

P = x..

= 0

Marginal distribution of X.

X	0	1	2
$P(X=x)$	$18/72$	$24/72$	$30/72$.

Marginal distribution of Y.

y	1	2	3
$P(y)$	$15/72$	$24/72$	$33/72$

Conditional distributions of X given Y.

To find:

c.d. of X given $Y=1$.

c.d. of X given $Y=2$.

c.d of X given $Y=3$.

Conditional dis of X given $Y=1$

X	$P(X=xi/Y=1)$.
0	$3K/15K = 3/15$.
1	$5K/15k = 5/15$
2	$7K/15K = 7/15$

$P(X=xi/Y=1)$

C.D of X given $Y=2$.

X	$P(X=xi/Y=2)$.
0	$6K/24K = 6/24$
1	$8K/24K = 8/24$
2	$10K/24K = 10/24$.

C.D of X given Y = 3.

X	$P(X = x_i / Y = 3)$.
0	$9K/33K = 9/33$
1	$24K/33K = 24/33$
2	$13K/33K = 13/33$.

Conditional distribulation of Y given X.

C.D of Y given X = 0.

Y	$P(Y = y_j / X = 0)$.
1	$3K/18K = 3/18$
2	$6K/18K = 6/18$.
3	$9K/18K = 9/18$

C.D of Y given X = 1.

Y	$P(Y = y_j / X = 1)$.
1	$5K/24K = 5/24$.
2	$8K/24K = 8/24$.
3	$11K/24K = 11/24$.

C.D. of Y given X = 2:

Y	$P(Y = y_j / X = 2)$.
1	$7K/30K = 7/30$.
2	$10K/30K = 10/30$
3	$13K/30K = 13/30$.

4] If X and Y have joint PDF,

$$f(x,y) = \begin{cases} x+y, & 0<x<1,\ 0<y<1. \\ 0, & \text{otherwise}. \end{cases}$$

check whether X and Y are independent.

solution,

To check independence of 2 variables.

$$f(x,y) = f(x) * f(y).$$

Marginal distribution of X.

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\, dy.$$

$$= \int_0^1 x+y \, dy.$$

$$= xy + y^2/2 \Big]_0^1$$

$$= x(1) + 1/2 + 0.$$

$$= x + 1/2.$$

$$= \frac{2x+1}{2}. \quad (\text{or}) \quad f(x) = x + 1/2.$$

Marginal distribution of Y.

$$f(y) = \int_{-\infty}^{\infty} f(x,y)\, dx.$$

$$= \int_0^1 x+y \, dx.$$

$$= \frac{x^2}{2} + xy \Big]_0^1$$

$$f(y) = \frac{1}{2} + y.$$

$$f(x,y) = f(x)\, f(y)$$

$$= (x + 1/2)(1/2 + y).$$

$$= \frac{x}{2} + xy + 1/4 + y/2.$$

$$= \frac{x}{2} + y/2 + xy + 1/4. \neq x + y.$$

$\therefore$ It is not independent.

5] If the joint PDF of a 2 dimensional variable is given by.

$$f(x,y) = K(6-x-y),\ 0 < x < 2,\ 2 < y < 4.$$

find:

(i) K (ii) $P(x<1, y<3)$ (iii) $P(x+y<3)$.

Solution:

Since, total probability $= 1$.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1.$$

$$\int_2^4 \int_0^2 K(6-x-y)\, dx\, dy$$

$$\int_2^4 K\left[6x - \frac{x^2}{2} - yx\right]_0^2 dy$$

$$\int_2^4 K[12 - 2 - 2y]\, dy.$$

$$\int_2^4 k[10-2y]\,dy = 1$$

$$k\left[10y - 2y^2/2\right]_2^4 = 1.$$

$$k\left[40 - \frac{32}{2} - 20 + 8/2\right] = 1.$$

$$k[40-16-20+4] = 1.$$

$$k[20-12] = 1.$$

$$k[8] = 1.$$

$$k = 1/8.$$

(ii) $P(X<1,\ Y<3)$:

$$\int_2^3 \int_0^1 f(x,y)\,dx\,dy.$$

$$= 1/8 \int_2^3 \int_0^1 6 - x - y\ dx\,dy.$$

$$= 1/8 \int_2^3 \left[6x - x^2/2 - yx\right]_0^1 dy.$$

$$= 1/8 \int_2^3 6 - 1/2 - y\,dy \Rightarrow 1/8 \int_2^3 6 - 1/2 - y\,dy.$$

$$= 1/8 \left[6y - y/2 - y^2/2\right]_2^3$$

$$= 1/8 \left[18 - 3/2 - 9/2 - 12 + 1 + 4/2\right]$$

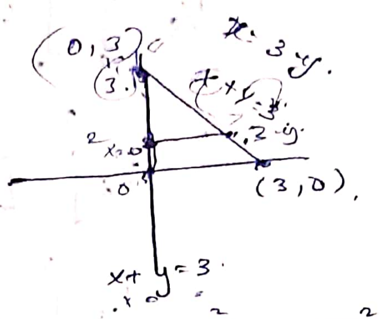
$$= 1/8\left[9 - \frac{12}{2}\right] \Rightarrow 1/8\left[\frac{18-12}{2}\right] \Rightarrow 1/8\left[\frac{6}{2}\right]^{3}$$

(

Scanned with OKEN Scanner

$P(x<1, y<3) = 3/8$.

(iii) $P(x+y<3)$.



$$= \frac{1}{8}\int_2^3\int_0^{3-y}(6-x-y)\,dx\,dy.$$

$$= \frac{1}{8}\int\left[6x - \frac{x^2}{2} - yx\right]_0^{3-y} dy.$$

$(3-y)^2 = 9-6y+y^2$

$$= \frac{1}{8}\int_2^3\left[6(3-y) - \frac{(3-y)^2}{2} - y(3-y)\right]dy.$$

$$= \frac{1}{8}\int_2^3\left[18-6y - \frac{9+6y-y^2}{2} - 3y+y^2\right]dy.$$

$$= \frac{1}{8}\int\left[\frac{36-12y-9+6y-y^2-6y+2y^2}{2}\right]dy.$$

$$= \frac{1}{8}\int_2^3\left[27-12y-y^2\right]$$

$(P(x+y<3)) = 5/24$.

(iv) $P(x<1 / y<3)$.

$$= \frac{P(x<1 \cap y<3)}{P(y<3)}.$$

$$= \frac{\int_2^3\int_0^1 f(x,y)\,dx\,dy}{\int_2^3\int_0^2 f(x,y)\,dx\,dy}.$$

$0<x<y<1$

$$= \frac{1}{8}\int_2^3\int_0^1 6-x-y\,dx\,dy.$$

$$= \frac{1}{8} \int_2^3 \int_0^1 6 - x - y \, dx \, dy.$$

$$= \frac{1}{8} \int_2^3 \left[6x - \frac{x^2}{2} - yx \right]_0^1 dy.$$

$6 - \frac{1}{2}$

$\frac{12-1}{2} \Rightarrow \frac{11}{2}.$

$$= \frac{1}{8} \int_2^3 6 - \frac{1}{2} - y \, dy.$$

$6 - \frac{1}{2}.$

$\frac{12-1}{2}.$

$\frac{11}{2}.$

$$= \frac{1}{8} \int_2^3 \frac{11}{2} - y \, dy.$$

$$= \frac{1}{8} \left[\frac{11y}{2} - \frac{y^2}{2} \right]_2^3$$

$$= \frac{1}{8} \left[\frac{33}{2} - \frac{9}{2} - \frac{22}{2} + \frac{4}{2} \right]$$

$$= \frac{1}{8} \left[\frac{33 - 9 - 22 + 4}{2} \right] =$$

$\frac{\cancel{6}^{3}}{\cancel{2}} \times \frac{1}{8}$ 4.

$\frac{\cancel{6}^{3}}{\cancel{2}} \times \frac{1}{8}$

$$= \frac{3}{8}.$$

$$= \frac{1}{8} \int_2^3 \int_0^2 6 - x - y \, dx \, dy.$$

$$= \frac{1}{8} \int_2^3 \left[6x - \frac{x^2}{2} - yx \right]_0^2 dy.$$

$$= \frac{1}{8} \int_2^3 \left[12 - \frac{4}{2} - 2y \right] dy.$$

$$= \frac{1}{8} \int_2^3 10 - 2y \, dy.$$

$$= \frac{1}{8} \left[10y - \frac{2y^2}{2} \right]_2^3$$

~~10(3)~~.

$$= \frac{1}{8} [30 - 9 - 20 + 4]$$

$= 5/8.$

$p(x<1/y<3) = \frac{3}{8} \times \frac{8}{5}$

$p(x<1/y<3) = 3/5.$

6] The joint PDF of random variable x and y is given by $f(x,y) = kxy\, e^{-(x^2+y^2)}$, $x>0, y>0$, find the value of k and also prove, that x and y are independent.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1.$$

$f(x) \cdot f(y) = f(x,y)$

$$k\int_{0}^{\infty}\int_{0}^{\infty} xy e^{-(x^2+y^2)}\,dx\,dy.$$

$$\Rightarrow \cancel{k\int_0^{\infty} x e^{-(x^2+y^2)}\,dy}.$$

$$= k\int_0^{\infty} x e^{-x^2}\,dx \cdot \int_0^{\infty} y e^{-y^2}\,dy.$$

Bernoulli's theorem on substitution

Let $\Rightarrow t = x^2 \Rightarrow x = \sqrt{t}$. $\frac{dx}{2x}$

$2x.$

$$= k\int_0^{\infty} \sqrt{t}\, e^{-t}\, \frac{dx}{2x} \cdot \int_0^{\infty}$$

$$= \int_0^{\infty} \sqrt{t}\, e^{-t}\, \frac{dt}{2\sqrt{t}} = \frac{1}{2}\int_0^{\infty} e^{-t}\,dt.$$

$= \frac{1}{2}\left(e^{-t}\right)_0^\infty$

$= \frac{1}{2}\left[0-(-1)\right]$

$= \frac{1}{2}$

$k\left[\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\right] = 1$

$\frac{K}{4} = 1$

$\boxed{K=4}$

To find : $f(x,y) = f(x) \cdot f(y)$

$f(x) = \int_0^\infty 4xy e^{-(x^2+y^2)} \, dy$

$= 4xe^{-x^2} \int_0^\infty y e^{-y^2} \, dy \qquad \because \int_0^\infty y e^{-y^2} \, dy = \frac{1}{2}$

$= 4xe^{-x^2}\left[\frac{1}{2}\right]$

$f(x) = 2xe^{-x^2}, \; x > 0$

Similarly

$f(y) = 2ye^{-y^2}, \; y > 0$

$f(x) \cdot f(y) = 2xe^{-x^2} \cdot 2ye^{-y^2}$

$f(x) \cdot f(y) = 4xy \, e^{-(x^2+y^2)}$

$f(x,y) = f(x) \cdot f(y)$

Independent.

Scanned with OKEN Scanner

A joint PDF of a 2 dimension random Variable X, Y is given by,

$$f(x,y) = xy^2 + x^2/8 \quad , \quad 0 \le x \le 2, \ 0 \le y \le 1.$$

Compute,

(i) $P(X > 1)$ (ii) $P(Y < 1/2)$ (iii) $P(X > 1 / Y > 1/2)$

(iv) $P(Y < 1/2 / X > 1)$ (v) $P(X < Y)$ (vi) $P(X + Y \le 1)$.

Solution :

given :

$$f(x,y) = xy^2 + x^2/8 .$$

(i) $P(X > 1)$:

$$= \int_1^2 f(x,y)\,dy . \Rightarrow \int_1^2 xy^2 + \frac{x^2}{8}\, dy.$$

$$\Rightarrow \left[xy^3/3 + \frac{x^2 y}{8} \right]_1^2 \Rightarrow \frac{8x}{3} + \frac{2x^2}{8} - \frac{x}{3} - \frac{x^2}{8} .$$

$$P(X>1) = \frac{7x}{3} + \frac{x^2}{8}$$

$$P(X>1) = \frac{x^2}{8} + \frac{7x}{3} .$$

$$= \int_0^1 \int_1^2 \left(xy^2 + \frac{x^2}{8} \right) dx\, dy .$$

$$= \int_0^1 \left[\frac{x^2 y^2}{2} + \frac{x^3}{24} \right]_1^2 dy .$$

$$= \int_0^1 \left[\frac{4y^2}{2} + \frac{8}{24} - \frac{y^2}{2} - \frac{1}{24} \right] dy .$$

$$\left[\frac{2y^3}{3} + y/3 - \frac{y^3}{6} - y/24 \right]_0^1$$

$$= \frac{2}{3} + 1/3 - 1/6 - 1/24 \Rightarrow \frac{16 + 8 - 4 - 1}{24} \Rightarrow \frac{19}{24}$$

(ii) $P(Y<1/2)$.

$$= \int_0^{1/2}\int_0^2 xy^2 + \frac{x^2}{8}\,dx\,dy.$$

$$= \int_0^{1/2}\left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_0^2 dy.$$

$$= \int_0^{1/2}\left[\frac{4y^2}{2} + \frac{8}{24}\right]dy.$$

$$= \int_0^{1/2} 2y^2 + 1/3\,dy \Rightarrow \left[\frac{2y^3}{3} + y/3\right]_0^{1/2}.$$

$$= \frac{2(\frac{1}{\cancel{8}\,4})}{3} + \frac{1}{6} \Rightarrow \frac{1}{12} + 1/6$$

$$= \frac{1+2}{12} \Rightarrow \frac{3}{12} \Rightarrow 1/4.$$

$P(Y<1/2) = 1/4$.

(iii) $P(X\geq 1 / Y<1/2)$.

$$\int_0^{1/2}\int_1^2 xy^2 + \frac{x^2}{8}\,dx\,dy.$$

$$= \int_0^{1/2}\left[\frac{x^2y}{2} + \frac{x^3}{24}\right]_1^2 dy.$$

$$= \int_0^{1/2}\left[\frac{4y^2}{2} + \frac{8}{24} - \frac{y}{2} + 1/24\right]dy.$$

$$= \int_0^{1/2}\left[2y^2 + 1/3 - y/2 + 1/24\right]dy$$

$$= \left[\frac{2y^3}{3} + y/3 - \frac{y^2}{4} + y/24\right]_0^{1/2}$$

Scanned with OKEN Scanner

$$\boxed{P(X>1 / Y<1/2) = \frac{P(X>1 \cap Y<1/2)}{P(Y<1/2)}.}$$

$$= \left[\frac{2(1/2)}{3} + 1/6 - 1/16 + 1/48\right].$$

$$= \frac{1}{12} + 1/6 - 1/16 + 1/48.$$

$$= \frac{4+8-3+1}{48} \Rightarrow \frac{13-3}{48}.$$

$$P(X>1 \cap Y<1/2) = \frac{\cancel{10}\,5}{\cancel{48}\,24} \Rightarrow \frac{5}{24}.$$

$$P(Y<1/2) = 1/4.$$

$$\therefore P(X>1 / Y<1/2) = 5/24 \times \frac{4}{1}$$

$$P(X>1/Y<1/2) = 5/6.$$

(iv) $P(Y<1/2 / X>1)$:

$$P(Y<1/2 / X>1) = \frac{P(Y<1/2 \cap X>1).}{P(X>1)}.$$

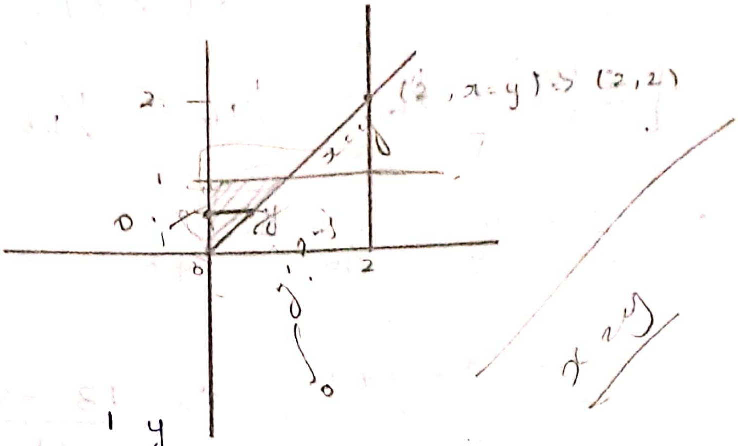
$$= \cancel{5/6 \times} \frac{5/24}{19/24}$$

$$= \cancel{5/24 \times 24/19} \Rightarrow \frac{5}{24} \times \frac{24}{19}.$$

$$= \cancel{\frac{5\times 4}{19}} \Rightarrow \cancel{\frac{20}{19}}.$$

$$P(Y<1/2 / X>1) = \frac{5}{19}.$$

(v) $P(X<Y)$. $\quad 0<x<2,\ 0<y<1.$



$$\int_0^1 \int_0^y xy^2 + \frac{x^2}{8}\, dx\, dy.$$

$$= \int_0^1 \left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_0^y dy.$$

$$\Rightarrow \int_0^1 \frac{y^2 y^2}{2} + \frac{y^3}{24}\, dy \Rightarrow \int_0^1 y^4/2 + y^3/24\, dy$$

$$= \left[y^5/10 + y^4/96 \right]_0^1$$

$$\Rightarrow \left[\frac{1}{10} + \frac{1}{96}\right] = \frac{96+10}{960} \Rightarrow \frac{\cancel{106}\ 53}{\cancel{960}\ 480}.$$

$$P(X<Y) = \frac{53}{480}.$$

(vi) $P(X+Y<1)$

2D discrete Random variable.

If the possible values (x, y) are finite or countably infinite, then (x, y) is called 2D discrete random variable.

2D continuous Random variables

If (x, y) can assume all values in a specified region R in the (x, y) plane.

Joint Probability Distributions of 2D random variable x and y.

We define the probability of the joint event $\{X \leq x, Y \leq y\}$, which is a function of a numbers x and y, by joint

Properties of the Joint Distribution:

* $F_{xy}(-\infty, \infty) = 0$
* $F_{xy}(-\infty, y) = 0$ and $F_{xy}(x, -\infty) = 0$.
* $0 \leq F_{xy}(x, y) \leq 1$
*

If x, y is a 2Dimensional discrete random variable, such that $P(X = x_i, Y = y_j) = P_{ij}$ is called the joint probability function (or) Joint Probability Mass Function of (x, y)

providing the following conditions are satisfied.

(i) $P_{ij} \geq 0$

(ii) $\sum_{i} \sum_{y} P_{ij} = 1$ (Total probability is equal to 1).

Marginal probability mass function of x.

If the joint probability distribution of two random variables x and y is given, then the marginal probability function of x is given by P.

Note:

Marginal Probability Mass Function of y:

If the joint probability distribution of two random variable x, and y is given then the marginal probability function of y is given by.

$$P_y(y_j) = P[y = y_j) =$$

NOTE:

The set $\{ y_j, P_{*j} \}$ is called the marginal distribution of y.

Conditional Probability Distribution:

CASE (i):

Probability of x_i given y_j is

$$P\{X = x_i / Y = y_j\} = \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}$$

CASE (ii):

The conditional probability of y_j given x_i is,

$$P\{Y = y_j / X = x_i\} = \frac{P[X = x_i, Y = y_j]}{P[X = x_i]}$$

Independent Random Variables

Two random variables x and y are said to be independent if.

* For discrete cases

$$\Rightarrow P_{ij} = P_{i*} \cdot P_{*j}$$

* For continuous Case:

$$\Rightarrow f(x, y) = f(x) \cdot f(y).$$

PROBLEMS:

7) The joint probability mass function of x, y is given by $P(x, y) = K(2x + 3y)$, $x = 0, 1, 2$, & $y = 1, 2, 3$. Find all the marginal and conditional distributions. Also find the marginal distribution $(x + y)$ and $P(x + y > 3)$.

Solution:

$P(x,y) = K(2x+3y)$ $x = 0,1,2$ & $y = 1,2,3$,

x \ y	1	2	3
0	3K $P(0,1)$	6K $P(0,2)$	9K $P(0,3)$
1	5K $P(1,1)$	8K $P(1,2)$	11K $P(1,3)$
2	7K $P(2,1)$	10K $P(2,2)$	13K $P(2,3)$

(i) To find K:

WKT: $\sum_i \sum_j P_{ij} = 1$

$3K + 6K + 9K + 5K + 8K + 11K + 7K + 10K + 13K = 1.$

$72K = 1$

$$\boxed{K = \frac{1}{72}}.$$

Sub $K = \frac{1}{72}$ in table:

	1	2	3	Marginal distribution of x.
0	$3/72$ $P(0,1)$	$6/72$ $P(0,2)$	$9/72$ $P(0,3)$	$P(X=0) = 18/72 = P_{0*}$ Sum of 1st row
1	$5/72$ $P(1,1)$	$8/72$ $P(1,2)$	$11/72$ $P(1,3)$	$P(X=1) = 24/72$ $= P_{1*}$
2	$7/72$ $P(2,1)$	$10/72$ $P(2,2)$	$13/72$ $P(2,3)$	$P(X=2) = 30/72 =$ $= P_{2*}$
Marginal distribution of y.	$P(Y=1) = 15/72$ P_{*1} (sum of 1st column)	$P(Y=2) = 24/72$ $= P_{*2}$	$P(Y=3) = 33/72$ $= P_{*3}$	1

Marginal distribution of x:

$P(X=0) = \frac{18}{72}$, $P(X=1) = \frac{24}{72}$, $P(X=2) = \frac{30}{72}$.

Marginal distribution of y:

$$P(y=1) = \frac{15}{72}, \quad P(y=2) = \frac{24}{72}, \quad P(y=3) = \frac{33}{72}.$$

The conditional distribution of x is given y:

(ie) $P(X = x_i / y = y_i)$

CASE (i): given $\boxed{y=1}$

$$P(x=0/y=1) = P(0,1)/P(y=1) = \frac{3/72}{15/72} = \frac{1}{5}$$

$$P(x=1/y=1) = P(1,1)/P(y=1) = \frac{5/72}{15/72} = 1/3$$

$$P(x=2/y=1) = P(2,1)/P(y=1) = \frac{7/72}{15/72} = 7/15$$

CASE (ii) $\boxed{y=2}$

$$P(x=0/y=2) = P(0,2)/P(y=2) = \frac{6/72}{24/72} = \frac{1}{4}$$

$$P(x=1/y=2) = P(1,2)/P(y=2) = \frac{8/72}{24/72} = \frac{1}{3}$$

$$P(x=2/y=2) = P(2,2)/P(y=2) = \frac{10/72}{24/72} = \frac{5}{12}.$$

CASE (iii) $\boxed{y=3}$

$$P(x=0/y=3) = P(0,3)/P(y=3) = \frac{9/72}{33/72} = 3/11$$

$$P(x=1/y=3) = P(1,3)/P(y=3) = \frac{11/72}{33/72} = 1/3.$$

$$P(x=2/y=3) = P(2,3)/P(y=3) = \frac{13/72}{33/72} = \frac{13}{33}$$

∵ always $P(x,y) \longrightarrow$ 1st x value, 2nd y value

Conditional Distribution of y given x is

$$P[y = y_i / x = x_i]$$

CASE (i) given $\boxed{x=0}$:

$P(y=1/x=0) = P(0,1) / P(x=0) = \dfrac{3/72}{18/72} = 1/6$

$P(y=2/x=0) = P(0,2) / P(x=0) = \dfrac{6/72}{18/72} = 1/3$

$P(y=3/x=0) = P(0,3) / P(x=0) = \dfrac{9/72}{18/72} = 1/2$

CASE (ii) given $\boxed{x=1}$:

$P(y=1/x=1) = P(1,1)/P(x=1) = \dfrac{5/72}{24/72} = 5/24$.

$P(y=2/x=1) = P(1,2)/P(x=1) = \dfrac{8/72}{24/72} = 1/3$.

$P(y=3/x=1) = P(1,3)/P(x=1) = \dfrac{11/72}{24/72} = 11/24$.

CASE (iii) given $\boxed{x=2}$:

$P(y=1/x=2) = P(2,1)/P(x=2) = \dfrac{7/72}{30/72} = 7/30$

$P(y=2/x=2) = P(2,2)/P(x=2) = \dfrac{10/72}{30/72} = 1/3$.

$P(y=3/x=2) = P(2,3)/P(x=2) = \dfrac{13/72}{30/72} = 13/30$.

Probability Distribution of $x+y$:

$x+y$		Probability
1	$P(0,1)$	$3/72$
2	$P(0,2)+P(1,1)$	$\frac{6}{72}+\frac{5}{72} = \frac{11}{72}$
3.	$P(0,3)+P(1,2)\ P(2,1)$	$\frac{9}{72}+\frac{8}{72}+\frac{7}{72} = \frac{24}{72}$

5 $P(2,3)$ $13/72$

Total 1

$$P(X+Y>3) = P(X+Y=4) + P(X+Y=5).$$

$$= \frac{21}{72} + \frac{13}{72} = \boxed{\frac{34}{72}}$$

(2). The 2D R.V. (x,y) has the JDF $f(x,y) = \frac{x+2y}{27}$ (ie) $\left(P(x,y) = \frac{x+2y}{27}\right)$. $x = 0,1,2$. $y = 0,1,2$. Find the conditional distribution of y given x and x given y and also find the marginal distribution function of x & y.

Sol:

x / y	0	1	2	marginal distribution of x
0	0 P(0,0)	2/27 P(0,1)	4/27 P(0,2)	$P(x=0) = 6/27$
1	1/27 P(1,0)	3/27 P(1,1)	5/27 P(1,2)	$P(x=1) = 9/27$
2	2/27 P(2,0)	4/27 P(2,1)	6/27 P(2,2)	$P(x=2) = 12/27$
Marginal distribution of y	P(y=0) 3/27	P(y=1) 9/27	P(y=2) 15/27	1

Marginal distribution of x:

$$P(X=0) = \frac{6}{27}, \quad P(X=1) = \frac{9}{27}, \quad P(X=2) = \frac{12}{27}.$$

marginal distribution of y:

$$P(y=0) = \frac{3}{27}, \quad P(y=1) = \frac{9}{27}, \quad P(y=2) = \frac{15}{27}.$$

The conditional distribution of x and y when x given y is

$$P(X = x_i / Y = y_i)$$

CASE (i): $y = 0$, $x = 0,1,2$

$$P(X=0/Y=0) = \frac{P(0,0)}{P(y=0)} = 0$$

$$P(X=1/Y=0) = \frac{P(1,0)}{P(y=0)} = \frac{1}{27} \times \frac{27}{3} = \frac{1}{3}$$

$$P(X=2/Y=0) = \frac{P(2,0)}{P(y=0)} = \frac{2}{27} \times \frac{27}{3} = \frac{2}{3}.$$

case (ii) $Y=1, \; X=0,1,2.$

$$P(X=0/Y=1) = \frac{P(0,1)}{P(y=1)} = \frac{2}{27} \times \frac{27}{9} = \frac{2}{9}$$

$$P(X=1/Y=1) = \frac{P(1,1)}{P(y=1)} = \frac{3}{27} \times \frac{27}{9} = \frac{1}{3}$$

$$P(X=2/Y=1) = \frac{P(2,1)}{P(Y=1)} = \frac{4}{27} \times \frac{27}{9} = \frac{4}{9}$$

case (iii) : $Y=2, \; X=0,1,2.$

$$P(X=0/Y=2) = \frac{P(0,2)}{P(y=2)} = \frac{4}{27} \times \frac{27}{15} = \frac{4}{15}$$

$$P(X=1/Y=2) = \frac{P(1,2)}{P(Y=2)} = \frac{5}{27} \times \frac{27}{15} = \frac{1}{3}$$

$$P(X=2/Y=2) = \frac{P(2,2)}{P(Y=2)} = \frac{6}{27} \times \frac{27}{15} \Rightarrow \frac{2}{5}$$

The conditional distribution of y given x is

$$P(Y=y_i / X=x_i) = \frac{P[x_i, y_j]}{P[x=x_i]}.$$

case (i) : $X=0.$

$$P(Y=0/X=0) = \frac{P(0,0)}{P(X=0)} = 0$$

$$P(Y=1/X=0) = \frac{P(0,1)}{P(X=0)} = \frac{2}{27} \times \frac{27}{6} = \frac{1}{3}$$

$$P(Y=2/X=0) = \frac{P(0,2)}{P(X=0)} = \frac{4}{27} \times \frac{27}{6} = \frac{2}{3}$$

case (ii) : $X=1$

$P(X=1/Y=0) = 1/9$, $P(X=1/Y=1) = 3/27 / 9/27 = 1/3$

$P(X=2/Y=2) = 5/27 / 9/27 = 5/9.$

case (iii) $X=2,$

$P(X=2/Y=0) = 2/27 / 12/27 = 1/6$

$P(X=2/Y=1) = 4/27 / 12/27 = 1/3.$

$P(X=2/Y=2) = 6/27 / 12/27 = 1/2$

Simple Linear regression model

$$y = a + bx + \epsilon$$ where the intercept a and slope b are constants and ϵ is random error. The parameters a and b are regression coefficients.

- ★ Least square method
- ★ Actual mean method
- ★ Assumed mean method
- ★ Properties.

MULTIPLE REGRESSION MODEL

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

REGRESSION EQUATION OF TWO VARIABLES

$$y = b_0 + b_1 x_1 + b_2 x_2$$

Normal equations are

$$\Sigma y = Nb_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2 \longrightarrow (1)$$

$$x_1 y = b_0 x_1 + b_1 x_1^2 + b_2 x_1 x_2$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 \longrightarrow (2)$$

$$x_2 y = b_0 x_2 + b_1 x_1 x_2 + b_2 x_2^2$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 \longrightarrow (3)$$

1 The following table gives no. of trails required to break a certain kind of forged alloy and the percentage of 2 alloys elements present in the metals.

No. of trails (y)	Percent of element A (x_1)	Percent of element B (x_2)
41	1	5
49	2	5
69	3	5
65	4	5
40	1	10
50	2	10
58	3	10
57	4	10
31	1	15
36	2	15
44	3	15
57	4	15
19	1	20
31	2	20
33	3	20
43 $\Sigma y = 723$	4 $\Sigma x_1 = 40$	20 $\Sigma x_2 = 200$

Fit a least square regression plane and use its equation to estimate the no of twist req to break one of the bars when $x_1 = 2.5$ $x_2 = 12$

Let the regression plane be

$$y = b_0 + b_1 x_1 + b_2 x_2$$

The normal equations are

$$\Sigma y = N b_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

x_1^2 $[\Sigma x_1^2]$	x_2^2 $[\Sigma x_2^2]$	$[\Sigma x_1 y]$ $x_1 y$	$[\Sigma x_2 y]$ $x_2 y$	$[\Sigma x_1 x_2]$ $x_1 x_2$
1	25	41	205	5
4	25	98	245	10
9	25	207	345	15
16	25	260	325	20
1	100	40	400	10
4	100	100	500	20
9	100	174	580	30
16	100	228	570	40
1	225	31	465	15
4	225	72	540	30
9	225	132	660	45
16	225	228	855	60
1	400	19	380	20
4	400	62	620	40
9	400	99	660	60
16	400	172	860	80
$\Sigma x_1^2 = 120$	$\Sigma x_2^2 = 3000$	$\Sigma x_1 y = 1963$	$\Sigma x_2 y = 8210$	$\Sigma x_1 x_2 = 500$

$$\Sigma y = N b_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$723 = N b_0 + 40 b_1 + 200 b_2$$

$$723 = 16 b_0 + 40 b_1 + 200 b_2 \longrightarrow (1)$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$1963 = 40 b_0 + 120 b_1 + 500 b_2 \longrightarrow (2)$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

$$8210 = 200 b_0 + 500 b_1 + 3000 b_2 \longrightarrow (3)$$

$b_0 = 46.43$

$b_1 = 7.77$

$b_2 = -1.65$

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$y = 46.43 + 7.77 x_1 - 1.65 x_2$$

When $x_1 = 2.5$, $x_2 = 12$

$$y = 46.43 + 7.77(2.5) - 1.65(12)$$

$$= 46.43 + 19.425 - 19.8$$

$$= 46.055$$

$$= 46.05$$

2) Processing variable and current gain of transistor in integrated circuit

Diffusion time (x_1)	1.5	2.5	0.5	1.2	2.6	0.3	2.4	2	0.7	1.6
Sheet resistance (x_2)	66	87	69	141	93	105	111	78	66	123
Current gain (y)	5.3	7.8	7.4	9.8	10.8	9.1	8.1	7.2	6.5	12.6

Fit a regression plane and estimate current gain when diffusion time is 2.2 hours and resistance is 90.

Let the regression plane be

$$y = b_0 + b_1 x_1 + b_2 x_2$$

The normal equations are

$$\Sigma y = N b_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

y	x_1	x_2	x_1^2	x_2^2	$x_1 x_2$	$x_1 y$	$x_2 y$
5.3	1.5	66	2.25	4356	99	7.95	349.8
7.8	2.5	87	6.25	7569	217.5	19.5	678.6
7.4	0.5	69	0.25	4761	34.5	3.7	510.6
9.8	1.2	141	1.44	19881	169.2	11.76	1381.6
10.8	2.6	93	6.76	8649	232.5	28.08	1004.4
9.1	0.3	105	0.09	11025	31.5	2.73	955.5
8.1	2.4	111	5.76	12321	266.4	19.44	899.1
7.2	2	78	4	6084	156	14.4	561.6
6.5	0.7	66	0.49	4356	46.2	4.55	429
12.6	1.6	123	2.56	15129	196.8	20.16	1549.8
$\Sigma y = 84.6$	$\Sigma x_1 = 15.3$	$\Sigma x_2 = 939$	$\Sigma x_1^2 = 29.85$	$\Sigma x_2^2 = 94131$	$\Sigma x_1 x_2 = 1458.9$	$\Sigma x_1 y = 132.27$	$\Sigma x_2 y = 8320.2$

$$\Sigma y = N b_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$84.6 = 10 b_0 + 15.3 b_1 + 939 b_2 \longrightarrow (1)$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$132.27 = 15.3 b_0 + 29.85 b_1 + 1458.9 b_2 \longrightarrow (2)$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

$$8320.2 = 939 b_0 + 1458.9 b_1 + 94131 b_2 \longrightarrow (3)$$

$$b_0 = 2.26$$
$$b_1 = 0.22$$
$$b_2 = 0.06$$

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$y = 2.26 + 0.22 x_1 + 0.06 x_2$$

When $x_1 = 2.2$ $x_2 = 90$

$$y = 2.26 + 0.22(2.2) + 0.06(90)$$
$$= 2.26 + 0.484 + 5.4$$
$$= 8.144$$
$$= 8.14$$

3) Rice yield $= x_1$ fertilizer $= x_2$ Rainfall $= x_3$

x_1	45	50	55	70	75	75	85
x_2	25	35	45	55	65	75	85
x_3	31	28	32	32	29	27	31

When $x_2 = 60$ and $x_3 = 25$

Take $x_1 = y$ $x_2 = x_1$ $x_3 = x_2$

$\therefore$ $x_1 = 60$
$x_2 = 25$

Let the regression plane be

$$y = b_0 + b_1 x_1 + b_2 x_2$$

The normal equations are

$$\Sigma y = N b_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
45	25	31	625	961	775	1125	1395
50	35	28	1225	784	980	1750	1400
55	45	32	2025	1024	1440	2475	1760
70	55	32	3025	1024	1760	3850	2240
75	65	29	4225	841	1885	4875	2175
75	75	27	5625	729	2025	5625	2025
85	85	31	7225	961	2635	7225	2635
$\Sigma y=455$	$\Sigma x_1=385$	$\Sigma x_2=210$	$\Sigma x_1^2=23975$	$\Sigma x_2^2=6324$	$\Sigma x_1x_2=11500$	$\Sigma x_1y=26925$	$\Sigma x_2y=13630$

$$\Sigma y = Nb_0 + b_1\Sigma x_1 + b_2\Sigma x_2$$

$$455 = 7b_0 + 385b_1 + 210b_2 \longrightarrow (1)$$

$$\Sigma x_1y = b_0\Sigma x_1 + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2$$

$$26925 = 385b_0 + 23975b_1 + 11500b_2 \longrightarrow (2)$$

$$\Sigma x_2y = b_0\Sigma x_2 + b_1\Sigma x_1x_2 + b_2\Sigma x_2^2$$

$$13630 = 210b_0 + 11500b_1 + 6324b_2 \longrightarrow (3)$$

$$b_0 = 9.00$$

$$b_1 = 0.68$$

$$b_2 = 0.60$$

$$y = b_0 + b_1x_1 + b_2x_2$$

$$y = 9 + 0.68x_1 + 0.60x_2$$

When $x_1=60$, $x_2=25$

$$y = 9 + 0.68(60) + 0.60(25)$$

$$y = 9 + 40.8 + 15$$

$$y = 64.8$$

LINEAR CORRELATION

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\,S_{yy}}}$$

$$S_{xy} = \Sigma xy - \frac{\Sigma x\Sigma y}{N}$$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{N}$$

1) Calculate r

X	6	2	10	4	8
Y	9	11	5	8	7

X	Y	XY	X^2	Y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\Sigma x=30$	$\Sigma y=40$	$\Sigma xy=214$	$\Sigma x^2=220$	$\Sigma y^2=340$

$$E(x) = \frac{\Sigma x}{N} = \frac{30}{5} = 6$$

$$E(y) = \frac{\Sigma y}{N} = \frac{40}{5} = 8$$

$$E(xy) = \frac{\Sigma xy}{N} = \frac{214}{5} = 42.8$$

$$E(x^2) = \frac{\Sigma x^2}{N} = \frac{220}{5} = 44$$

$$E(y^2) = \frac{\Sigma y^2}{N} = \frac{340}{5} = 68$$

$$S_{xy} = 214 - \frac{30 \times 40}{5} = -26$$

$$S_{xx} = 220 - \frac{30^2}{5} = 40$$

$$S_{yy} = 340 - \frac{40^2}{5} = 20$$

$$r = \frac{-26}{\sqrt{40 \times 20}} = -0.919 \text{ or } (-0.921)$$

MULTIPLE CORRELATION

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2(r_{12} r_{13} r_{23})}{1 - r_{23}^2}}$$

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2(r_{21} r_{23} r_{13})}{1 - r_{13}^2}}$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2(r_{31} r_{12} r_{23})}{1 - r_{12}^2}} \text{ or } r_{32}$$

1) From the following data obtain multiple correlations $R_{1.23}$, $R_{2.13}$, $R_{3.12}$

x_1	2	5	7	11
x_2	3	6	10	12
x_3	1	3	6	10

x_1	x_2	$x_1 x_2$	x_1^2	x_2^2
2	3	6	4	9
5	6	30	25	36
7	10	70	49	100
11	12	132	121	144
$\Sigma x_1 = 25$	$\Sigma x_2 = 31$	$\Sigma x_1 x_2 = 238$	$\Sigma x_1^2 = 199$	$\Sigma x_2^2 = 289$

$$E(X_1) = \frac{\Sigma X_1}{N} = \frac{25}{4} = 6.25$$

$$E(X_2) = \frac{\Sigma X_2}{N} = \frac{31}{4} = 7.75$$

$$E(X_1 X_2) = \frac{\Sigma X_1 X_2}{N} = \frac{238}{4} = 59.5$$

$$E(X_1^2) = \frac{\Sigma X_1^2}{N} = \frac{199}{4} = 49.75$$

$$E(X_2^2) = \frac{\Sigma X_2^2}{N} = \frac{289}{4} = 72.25$$

$$\sigma_{x_1} = \sqrt{E(X_1^2) - [E(X_1)]^2} = \sqrt{49.75 - (6.25)^2} = \sqrt{49.75 - 39.0625} = \sqrt{10.6875} = 3.26$$

$$\sigma_{x_2} = \sqrt{E(X_2^2) - [E(X_2)]^2} = \sqrt{72.25 - (7.75)^2} = \sqrt{72.25 - 60.0625} = \sqrt{12.1875} = 3.49$$

$$r_{12} = \frac{E(X_1 X_2) - E(X_1).E(X_2)}{\sigma_{x_1} \sigma_{x_2}}$$

$$= \frac{59.5 - (6.25)(7.75)}{3.26 \times 3.49}$$

$$= \frac{59.5 - 48.4375}{11.3774}$$

$$= \frac{11.0625}{11.3774}$$

$$r_{12} = 0.97$$

x_2	x_3	$x_2 x_3$	x_2^2	x_3^2
3	1	3	9	1
6	3	18	36	9
10	6	60	100	36
12	10	120	144	100
$\Sigma x_2 = 31$	$\Sigma x_3 = 20$	$\Sigma x_2 x_3 = 201$	$\Sigma x_2^2 = 289$	$\Sigma x_3^2 = 146$

$$S_{xy} = S_{x_2 x_3} = \Sigma x_2 x_3 - \frac{\Sigma x_2 \Sigma x_3}{N} = 201 - \frac{(20)(31)}{4}$$

$$S_{xx} = S_{x_2 x_2} = \Sigma x_2^2 - \frac{(\Sigma x_2)^2}{N} = 289 - \frac{(31)^2}{4}$$

$$S_{yy} = S_{x_3 x_3} = \Sigma x_3^2 - \frac{(\Sigma x_3)^2}{N} = 146 - \frac{(20)^2}{4}$$

$$S_{x_2x_3} = 201 - \frac{(20)(31)}{4}$$

$$= 201 - \frac{620}{4}$$

$$= 201 - 155$$

$$= 46$$

$$r_{23} = \frac{S_{x_2x_3}}{\sqrt{S_{x_2x_2} S_{x_3x_3}}}$$

$$= \frac{46}{\sqrt{48.75 \cdot 46}}$$

$$= 0.97$$

$$S_{x_2x_2} = 289 - \frac{(31)^2}{4}$$

$$= 48.75$$

$$S_{x_3x_3} = 146 - \frac{(20)^2}{4}$$

$$= 46$$

x_1	x_3	$x_1 x_3$	x_1^2	x_3^2
2	1	2	4	1
5	3	15	25	9
7	6	42	49	36
11	10	110	121	100
$\Sigma x_1 = 25$	$\Sigma x_3 = 20$	$\Sigma x_1x_3 = 169$	$\Sigma x_1^2 = 199$	$\Sigma x_3^2 = 146$

$$S_{x_1x_3} = 169 - \frac{(25)(20)}{4} = 44$$

$$S_{x_1x_1} = 199 - \frac{(25)^2}{4} = 42.75$$

$$S_{x_3x_3} = 146 - \frac{(20)^2}{4} = 46$$

$$r_{13} = \frac{S_{x_1x_3}}{\sqrt{S_{x_1x_1} S_{x_3x_3}}} = \frac{44}{\sqrt{(42.75)(46)}} = 0.99$$

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2(r_{12} r_{13} r_{23})}{1 - r_{23}^2}} = 0.99$$

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2(r_{21} r_{23} r_{13})}{1 - r_{13}^2}} = 0.97$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2(r_{31} r_{32} r_{12})}{1 - (r_{12})^2}} = 0.99$$

2 Obtain $R_{1.23}$ $R_{2.13}$

x_1	65	72	54	68	55	59	78	58	57	51
x_2	56	58	48	61	50	51	55	48	52	42
x_3	9	11	8	13	10	8	11	10	11	7

$r_{12} = 0.799$

$r_{13} = 0.635$

$r_{23} = 0.795$

x_1	x_2	$x_1 x_2$	x_1^2	x_2^2
65	56	3640	4225	3136
72	58	4176	5184	3364
54	48	2592	2916	2304
68	61	4148	4624	3721
55	50	2750	3025	2500
59	51	3009	3481	2601
78	55	4290	6084	3025
58	48	2784	3364	2304
57	52	2964	3249	2704
51	42	2142	2601	1764
$\Sigma x_1 = 617$	$\Sigma x_2 = 521$	$\Sigma x_1x_2 = 32495$	$\Sigma x_1^2 = 38753$	$\Sigma x_2^2 = 27423$

$$S_{x_1x_2} = 32495 - \frac{(617)(521)}{10} = 349.3$$

$$S_{x_1x_1} = 38753 - \frac{(617)^2}{10} = 684.1$$

$$S_{x_2x_2} = 27423 - \frac{(521)^2}{10} = 278.9$$

$$\boxed{r_{12} = \frac{S_{x_1} S_{x_2}}{\sqrt{S_{x_1x_1} S_{x_2x_2}}} = \frac{349.3}{\sqrt{(684.1)(278.9)}} = 0.799}$$

x_2	x_3	$x_2 x_3$	x_2^2	x_3^2
56	9	504	3136	81
58	11	638	3364	121
48	8	384	2304	64
61	13	793	3721	169
50	10	500	2500	100
51	8	408	2601	64
55	11	605	3025	121
48	10	480	2304	100
52	11	572	2704	121
42	7	294	1764	49
$\Sigma x_2 = 473$	$\Sigma x_3 = 98$	$\Sigma x_2x_3 = 5178$	$\Sigma x_2^2 = 27423$	$\Sigma x_3^2 = 990$

$$S_{x_2x_3} = 5178 - \frac{(473)(98)}{10} = 542.6$$

$$S_{x_2x_2} = 27423 - \frac{(473)^2}{10} = 5050.1$$

$$S_{x_3x_3} = 990 - \frac{(98)^2}{10} = 29.6$$

$$\boxed{r_{23} = \frac{542.6}{\sqrt{(5050.1)(29.6)}} = 0.795}$$

X_1	X_3	X_1X_3	X_1^2	X_3^2
65	9	585	4225	81
72	11	792	5184	121
54	8	432	2916	64
68	13	884	4624	169
55	10	550	3025	100
59	8	472	3481	64
78	11	858	6084	121
58	10	580	3364	100
57	11	627	3249	121
51	7	357	2601	49
$\Sigma X_1=617$	$\Sigma X_3=98$	$\Sigma X_1X_3=6137$	$\Sigma X_1^2=38753$	$\Sigma X_3^2=990$

$$S_{x_1x_3} = 6137 - \frac{(617)(98)}{10} = 90.4$$

$$S_{x_1x_1} = 38753 - \frac{(617)^2}{10} = 684.1$$

$$S_{x_3x_3} = 990 - \frac{(98)^2}{10} = 29.6$$

$$\boxed{r_{13} = \frac{90.4}{\sqrt{(684.1)(29.6)}} = 0.635}$$

$$R_{1.23} = \sqrt{\frac{(0.799)^2+(0.635)^2-2(0.799)(0.635)(0.795)}{1-(0.795)^2}}$$

$$= 0.799$$

$$R_{2.13} = \sqrt{\frac{(0.799)^2+(0.795)^2-2(0.799)(0.635)(0.795)}{1-(0.635)^2}}$$

$$= 0.777$$

0.88

PARTIAL CORRELATION

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

$$r_{23.1} = \frac{r_{23} - r_{21}r^3{}_1}{\sqrt{(1-r_{21}^2)(1-r_{31}^2)}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}}$$

1) From the data relating to yield of dry bark (x_1) height (x_2) and girth (x_3) for 18 cinchonna plant following correlation coefficient were obtained. $r_{12}=0.77$, $r_{13}=0.72$, $r_{23}=0.52$. Find $r_{12.3}$ and $r_{23.1}$

Soln:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23})^2}}$$

$$= \frac{0.77-(0.72)(0.52)}{\sqrt{(1-(0.72)^2)(1-(0.52)^2}}$$

$$= 0.68$$

$$r_{23.1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{(1-r_{21}^2)(1-r_{31}^2)}}$$

$$= \frac{0.52-(0.77)(0.72)}{\sqrt{(1-(0.77)^2)(1-(0.72)^2)}}$$

$$= -0.08$$

2 In a bivariate distribution $\sigma_1 = 2$ $\sigma_2 = \sigma_3 = 3$, $r_{12} = 0.7$ $r_{23} = r_{31} = 0.5$. Find

i) $R_{1.23}$

ii) $r_{23.1}$

iii) $b_{12.3}$ and $b_{13.2}$

iv)

Soln:

i) $$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2(r_{12}r_{13}r_{23})}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.7)^2 + (0.5)^2 - 2(0.7 \times 0.5 \times 0.5)}{1 - (0.5)^2}}$$

$$= \sqrt{\frac{0.49 + 0.25 - 2(0.175)}{1 - 0.25}}$$

$$= \sqrt{\frac{0.49 + 0.25 - 0.35}{0.75}}$$

$$= \sqrt{\frac{0.39}{0.75}}$$

$$= \sqrt{0.52}$$

$$= 0.72$$

ii) $$r_{23.1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{(1 - r_{21}^2)(1 - r_{31}^2)}}$$

$$= \frac{0.5 - 0.7 \times 0.5}{\sqrt{(1 - (0.7)^2)(1 - (0.5)^2)}}$$

$$= \frac{0.5 - 0.35}{\sqrt{(1 - 0.49)(1 - 0.25)}}$$

$$= \frac{0.15}{\sqrt{(0.51)(0.75)}}$$

$$= \frac{0.15}{\sqrt{0.3825}}$$

$$= \frac{0.15}{0.61}$$

$$= 0.24$$

iii) $b_{12.3}$ and $b_{13.2}$

$$b_{xy} = \rho \frac{\sigma_x}{\sigma_y}$$

$$b_{12.3} = \frac{r_{12.3}\sigma_{1.3}}{\sigma_{2.3}}$$

$\sigma_{1.3} = \sigma_1\sqrt{1 - r_{13}^2}$

$\sigma_{2.3} = \sigma_2\sqrt{1 - r_{23}^2}$

$\sigma_{1.3} = 2\sqrt{1 - (0.5)^2} = 2\sqrt{1 - 0.25} = 2(0.75) = 1.5$

$\sigma_{2.3} = 3\sqrt{1 - (0.5)^2} = 3\sqrt{1 - 0.25} = 3(0.75) = 2.25$

$$b_{12.3} = \frac{r_{12.3} \times 1.5}{2.25}$$

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13})^2(1 - r_{23})^2}}$$

$$= \frac{0.7 - (0.5)(0.5)}{\sqrt{(1 - (0.5)^2)(1 - (0.5))^2}}$$

$$= \frac{0.7 - (0.5)^2}{\sqrt{(1 - 0.25)(1 - 0.25)}}$$

$$= \frac{0.7 - (0.5)^2}{\sqrt{(0.75)(0.75)}}$$

$$= \frac{0.7 - (0.5)^2}{\sqrt{(0.75)^2}} = 0.6$$

$$b_{12.3} = \frac{0.6 \times 1.5}{2.25}$$

$$= 0.4$$

$$b_{13.2} = \frac{r_{13.2}\,\sigma_{1.2}}{\sigma_{3.2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12} r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}}$$

$$= \frac{0.5 - 0.7 \times 0.5}{\sqrt{(1-(0.7)^2)(1-(0.5)^2)}}$$

$$= \frac{0.5 - 0.35}{\sqrt{(1-0.49)(1-0.25)}}$$

$$= \frac{0.15}{\sqrt{(0.51)(0.75)}}$$

$$= \frac{0.15}{\sqrt{0.3825}}$$

$$= 0.24$$

$$\sigma_{1.2} = \sigma_1\sqrt{1-r_{12}^2}$$

$$= 2\sqrt{1-(0.7)^2}$$

$$= 2\sqrt{1-0.49}$$

$$= 2\sqrt{0.51}$$

$$= 1.42$$

$$\sigma_{3.2} = \sigma_3\sqrt{1-r_{32}^2}$$

$$= 3\sqrt{1-(0.5)^2}$$

$$= 3\sqrt{1-0.25}$$

$$= 3\sqrt{0.75}$$

$$= 2.59$$

$$b_{13.2} = \frac{r_{13.2}\,\sigma_{1.2}}{\sigma_{3.2}}$$

$$= \frac{0.24 \times 1.42}{2.59}$$

$$= 0.13$$

Properties

1) Multiple correlation coefficient is the degree of observed value of the dependent variable and its estimate obtained by multiple regression

2) Multiple correlation coefficient lies between 0 and 1

3) If multiple correlation coefficient is 1 then association is perfect and multiple regression equation may said to be perfect prediction formula.

4) If multiple correlation coefficient is 0 dependent variable is uncorrelated with other independent variables. From this it can be concluded that multiple regression equation fails to predict the value of dependent variable when values of independent variable are known.

5) Multiple correlation coefficient is always greater than or equal to any total correlation coefficient (i.e.)

$$R_{1.23} \geq r_{12} \text{ or } r_{13} \text{ or } r_{23}$$

6) Multiple correlation coefficient obtained by the method of least squares would always be greater than the multiple correlation coefficient obtained by any other method

7) Correlation coefficient is geometric mean of regression coefficient.

$$r_{12.3} = b_{12.3} \times b_{21.3}$$

1) Find a regression equation of X_1 (acre per seed) on X_2 (rainfall in inches) and X_3 (accumulated temperature above 42°c) from the following data

TRAIT	MEAN	SD	r_{12}	r_{23}	r_{13}
x_1	28.02	4.42	0.8	-	-
x_2	4.91	1.10	-	-0.56	-
x_3	594	85	-	-	-0.40

$$(X_1-\bar{X}_1)\frac{\omega_{11}}{\sigma_1} + (X_2-\bar{X}_2)\frac{\omega_{12}}{\sigma_2} + (X_3-\bar{X}_3)\frac{\omega_{13}}{\sigma_3} = 0$$

$$\omega = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0.8 & -0.4 \\ 0.8 & 1 & -0.56 \\ -0.4 & -0.56 & 1 \end{vmatrix}$$

$$\omega_{11} = \begin{vmatrix} 1 & -0.56 \\ -0.56 & 1 \end{vmatrix} = 0.69$$

$$\omega_{12} = \begin{vmatrix} 0.8 & -0.56 \\ -0.4 & 1 \end{vmatrix} = 0.576 = 0.58$$

$$\omega_{13} = \begin{vmatrix} 0.8 & 1 \\ -0.4 & -0.56 \end{vmatrix} = -0.048 = -0.05$$

The regression equation is

$$(X_1 - 28.02)\frac{0.69}{4.42} + (X_2 - 4.91)\frac{0.58}{1.10} + (X_3 - 594)\frac{(-0.05)}{85} = 0$$

$$(X_1 - 28.02)(0.156) + (X_2 - 4.91)(0.527) + (X_3 - 594)(0.0006) = 0$$

$$0.156X_1 + 0.53X_2 - 0.0006X_3 = 6.6$$

$\div\ 0.156)$

$$x_1 = 42.3 - 3.4X_2 + 0.0006X_3$$

2) The following constants are obtained from measurement on length in mm (X_1), volume in cc (X_2) and weight in gm (X_3) of 300 eggs.

	MEAN	SD	r_{12}	r_{23}	r_{31}
x_1	55.95	2.26	0.578	–	–
x_2	51.48	4.39	–	0.974	–
x_3	56.03	4.41	–	–	0.581

Obtain the linear regression equation of egg weight on egg length and egg volume. Hence estimate the weight of an egg whose length is 58 cm and volume is 52.5 cc

Soln:

Reg eqn of X_3 on X_1 and X_2

$$(X_3-\bar{X}_3)\frac{\omega_{33}}{\sigma_3} + (X_2-\bar{X}_2)\frac{\omega_{32}}{\sigma_2} + (X_1-\bar{X}_1)\frac{\omega_{31}}{\sigma_1} = 0$$

$$\omega = \begin{vmatrix} 1 & 0.578 & 0.581 \\ 0.578 & 1 & 0.974 \\ 0.581 & 0.974 & 1 \end{vmatrix}$$

$$\omega_{32} = \begin{vmatrix} 0.578 & 0.581 \\ 1 & 0.974 \end{vmatrix} = -0.01$$

$$\omega_{32} = \begin{vmatrix} 1 & 0.581 \\ 0.578 & 0.974 \end{vmatrix} = 0.64\ /\ 0.638$$

$$\omega_{33} = \begin{vmatrix} 1 & 0.578 \\ 0.578 & 1 \end{vmatrix} = 0.67\ /\ 0.666$$

The regression eqn is,

$$(X_1 - 55.95)\frac{-0.018}{2.26} + (X_2 - 51.48)\frac{0.638}{4.39} + (X_3 - 56.03)\frac{0.666}{4.41} = 0$$

$$-0.007X_1 + 0.145X_2 + 0.151X_3 = 15.534$$

$\div$ by 0.151

$$X_3 = 0.05X_1 - 0.97X_2 + 102.9$$

When $X_1 = 58$ $X_2 = 52.5$

$$X_3 = 0.05(58) - 0.97(52.5) + 102.9$$

$$= 55.12$$

ANALYSIS OF VARIANCE (ANOVA)

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

ASSUMPTIONS IN ANALYSIS OF VARIANCE

1.) Each of the samples is drawn from a normal population

2.) The variances for the population from which samples have been drawn are equal.

3.) The variation of each value around the own grand mean should be independent for each value.

BASIC STEPS IN ANALYSIS OF VARIANCE

Determine

1) One estimate of the population variance from the variance among the sample means.

2) Determine a second estimate of the population variance from the variance within the sample mean.

3) Compare these two estimates if they are approximately equal in value, ~~en~~ accept the null hypothesis.

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F RATIO
BETWEEN SAMPLES	SSC	$\gamma_1 = k-1$	$MSC = \frac{SSC}{K-1}$	$F = \frac{MSC}{MSE}$
WITHIN SAMPLES	SSE	$\gamma_2 = N-k$	$MSE = \frac{SSE}{N-E}$	

1) A common test was given to a number of students taken at random from a particular class of 4 departments concerned to access the significance of possible variation in performance. Make an analysis of variance given from the table.

DEPARTMENTS			
C	M	E	I
9	12	17	13
10	13	17	12
13	11	15	12
9	14	9	18
9	5	7	15

NEYMAN - PEARSON LEMMA :

Suppose we have a random sample $x_1, x_2, \ldots, x_n$ from a probability distribution with parameter θ. Then if C is a critical region of size α and k is a constant such that

$$\frac{L(\theta_0)}{L(\theta_\alpha)} \le k \text{ inside the critical region } C$$

and

$$\frac{L(\theta_0)}{L(\theta_\alpha)} \ge k \text{ outside the critical region } C$$

then C is the best, most powerful, critical region for testing the simple null hypothesis $H_0 : \theta = \theta_0$ against the sample alternative hypothesis $H_A : \theta = \theta_\alpha$.

MODULE – II

TESTING OF HYPOTHESIS

TESTS OF SIGNIFICANCE:

Tests of significance, which enable us to decide on the basis of the sample results in

(i) The deviation between the observed sample statistic and the hypothetical parameter value is significant.

(ii) The deviation between two sample statistic is significant.

NULL HYPOTHESIS (H_0):

A defenite statement about the population parameter (hypothesis of no difference)

ALTERNATIVE HYPOTHESIS (H_1):

Any hypothesis which is complementary to Null hypothesis is called an alternative hypothesis(H

Eg: $H_0 = 50$

$H_1 \neq 50$ [$H_1 < 50$, or $H_1 > 50$]
(Left tailed) (Right tailed)

ERRORS IN SAMPLING:

	ACCEPTED	REJECTED
H_0 TRUE	✓	× Type I error
H_0 FALSE	× Type II error	✓

TYPE 1 ERROR (α) Rejection of H_0, when H_0 is true

TYPE 2 ERROR (β) Acceptance of H_0, when H_0 is false

$\alpha \rightarrow$ Producers Risk $\qquad$ $\beta \rightarrow$ Consumer's Risk

PROCEDURE FOR TESTING OF HYPOTHESIS:

1. Set up the null hypothesis.
2. Choose the appropriate level of significance (5% or 1%)
3. Compute the test statistic, $z = \dfrac{t - E(t)}{SE(t)}$
4. Compare the computed value of z with the significant value at given level of significance

(i) If $|z| < 1.96$, H_0 may be accepted at 5% level

If $|z| > 1.96$, H_0 may be rejected at 5% level

(ii) If $|z| < 2.58$, H_0 may be accepted at 1% level

If $|z| > 2.58$, H_0 may be rejected at 1% level

CRITICAL REGION OR REJECTION REGION:

A region corresponding to a statistics (t) in the sample space is', which leads to rejection of H_0, is called critical region.

ACCEPTANCE REGION:

A region which leads to acceptance of H_0, is called acceptance region

LEVEL OF SIGNIFICANCE:

The probability α that a random value of statistic belongs to the critical region is known as level of significance.

TESTS OF SIGNIFICANCE OF SMALL SAMPLES:

When the size of sample is less than 30

It is called small sample.

1. Students t test

2. F test

3. χ^2 test (chi square test)

STUDENT'S TEST:

1. Students t test for single mean:

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

$\bar{x} \rightarrow$ sample mean

$\mu \rightarrow$ population mean

with dof. $\nu = n-1$

2. Students t test for difference of mean:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ with dof $n_1 + n_2 - 2$

$$S^2 = \frac{\Sigma (x_1 - \bar{x}_1)^2 + \Sigma (x_2 - \bar{x}_1)^2}{n_1 + n_2 - 2}$$

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

1. A sample of 26 bulbs gives a mean life of 990 hours with a SD of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not up to the standard?

NULL HYPOTHESIS (H_0):

The sample upto the standard.

Level of significance: 5%.

$n = 26$, $\bar{x} = 990$, $S = 20$, $\mu = 1000$

Test statistic: $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$

$$t = \frac{990 - 1000}{\frac{20}{\sqrt{26}}} = -2.55$$

$|t|_{cal} = 2.55$

$d.f = n - 1 = 26 - 1 = 25$

$t_{tab} = 1.708$

$t_{cal} > t_{tab}$

$\therefore H_0$ is rejected

The sample is not upto the standard.

---

2. A machinist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a SD of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specification.

H_0: NULL HYPOTHESIS:

The work is meeting the specification.

Level of significance: 5%.

$n = 10$, $\bar{X} = 0.742$, $S = 0.040$, $\mu = 0.700$

Test statistic: $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$

$$t = \frac{0.742 - 0.700}{\frac{0.040}{\sqrt{10}}} = 3.32$$

$t_{cal} = 3.32$

$d.f = n-1 = 10-1 = 9$

$t_{tab} = 1.833$

$t_{cal} > t_{tab}$

$\therefore H_0$ is rejected

The work is not meeting the specification.

3\. A random sample of size 16 values from a normal population, showed a mean of 53 and a sum of squares of deviation from the mean equals to 150. Can this sample be regarded as taken from the population, having 56 as a mean, obtain 95% confidence limits of the mean of the population.

$n = 16$, $\bar{x} = 53$ $\quad \sum (x-\bar{x})^2 = 150 \quad \mu = 56$

$$S = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{150}{15}} = 3.16.$$

H_0 : NULL HYPOTHESIS:

Sample taken from the population, having 56 as mean.

Test statistics:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$t = \frac{53-56}{\frac{3.16}{\sqrt{16}}} = -3.79$$

$cal_{|t|} = 3.8$

$t_{tab} = 1.753$

$t_{cal} > t_{tab}$

$\therefore H_0$ is rejected

Sample is not taken from the population, having 56 as mean.

Confidence limits : $\bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}$

$$53 \pm (1.753) \frac{\sqrt{10}}{\sqrt{16}}$$

$$= 53 - 1.753 \sqrt{\frac{10}{16}} \text{ to } 53 + 1.753 \sqrt{\frac{10}{16}}$$

$$= (51.61, 54.38)$$

4. The 9 items of a sample had the following values,

45, 47, 50, 52, 48, 47, 49, 53, 51.

Does the mean of 9 items differ significantly from the assumed population mean 47.5?

NULL HYPOTHESIS : H_0 :

The mean of 9 items does not differ significantly from the assumed population mean 47.5.

LOS = 5%

Given : $n = 9$, $\bar{x} = ?$, $\mu = 47.5$, $S = ?$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{45 + \dots + 51}{9} = 49.1$$

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

x	45	47	50	52	48	47	49	53	51
$(x-\bar{x})^2$	16.8	4.4	0.81	8.41	1.21	4.41	0.01	15.2	3.61

$$S = \sqrt{\frac{54.88}{9-1}} = 2.619.$$

TEST STATISTICS:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{49.1 - 47.5}{\frac{2.62}{\sqrt{9}}} = 1.832$$

$$|t|_{cal} = 1.832$$

$$d.f = \nu = n - 1 = 8$$

$$|t|_{tab} = 1.86$$

$$t_{cal} < t_{tab}$$

$\therefore H_0$ is accepted

The mean of 9 items does not differ significantly from the assumed population mean 47.5.

5. A random sample of 10 boys had the following IQ's:-

70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

Do these data support the assumptions of a population mean IQ of 100. Find a reasonable range in which most of the mean IQ is values of samples of 10 boys.

NULL HYPOTHESIS H_0:

Data supports the assumptions of a population mean IQ of 100.

LOS = 5%.

$n = 10$, $\bar{x} = ?$, $\mu = 100$.

$$\bar{x} = \frac{\sum x}{n} = 97.2$$

x	70	120	110	101	88	83	95	98	107	10
$(x-\bar{x})^2$	739.8	519.8	163.8	14.4	84.6	201.6	4.84	0.64	96.04	7

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{1833.36}{9}} = 14.27$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{97.2 - 100}{\frac{14.27}{\sqrt{10}}} = \frac{-2.8}{4.5} = -0.62$$

$|t|_{cal} = 0.62$

$d.f = \nu = n-1 = 9$

$|t|_{tab} = 1.833$

$t_{cal} < t_{tab}$

$\therefore H_0$ is accepted.

$\therefore$ Data supports the assumptions of a population mean IQ of 100.

Confidence limits / fiducial limits:

$$= \bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}$$

$$= 97.2 \pm (1.833) \frac{14.27}{\sqrt{10}}$$

$$= 97.2 \pm 8.278$$

$$= (97.2 - 8.278), (97.2 + 8.278)$$

$$= (88.92, 105.48)$$

STUDENTS T TEST FOR DIFFERENCE OF MEANS:

1. Samples of 2 types of electric bulbs were testested for length of like, and the following data were obtained.

	TYPE-I	TYPE-II
SAMPLE No.	$n_1 = 8$	$n_2 = 7$
SAMPLE MEANS	$\bar{x}_1 = 1234$ hrs	$\bar{x}_2 = 1036$ hrs
SAMPLE SD	$S_1 = 36$ hrs	$S_2 = 40$ hrs

Is the difference in the means sufficient to warrant that type I is superior to type II, regarding length of life?

NULL HYPOTHESIS : H_0:

Length of life of Type I and Type II are same.

LOS = 5%.

TEST STATISTIC:

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

$$= \frac{8(36)^2 + 7(40)^2}{8+7-2} = \frac{21568}{13}$$

$$S^2 = 1659.08$$

$$S = 40.73$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1234 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}}$$

$$= \frac{198}{40.73(0.518)} = \frac{198}{21.1} = 9.4$$

$$|t|_{cal} = 9.4$$

$$d.f = n_1 + n_2 - 2 = 13$$

$$|t|_{tab} = 1.771$$

$$t_{cal} > t_{tab}$$

$\therefore$ H_0 is rejected.

Type I is superior to type II.

2. The means of 2 random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sum of squares of deviation from the mean are 26.94 and 18.73. Can the sample be considered to have been drawn from the same normal population.

H_0: The samples drawn from the same normal population

$\bar{x}_1 = 196.42$ $\quad$ $\bar{x}_2 = 198.82$ $\quad$ $\sum (x_1-\bar{x}_1)^2 = 26.94$

$n_1 = 9$ $\quad$ $n_2 = 7$ $\quad$ $\sum (x_2-\bar{x}_2)^2 = 18.73$

$$S = \sqrt{\frac{\sum (x_1-\bar{x}_2)^2 + \sum (x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{26.94 + 18.73}{9+7-2}} = 1.81$$

Test statistics:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{196.42 - 198.82}{1.81\sqrt{\frac{1}{9}+\frac{1}{7}}} = -2.63$$

$|t|_c = 2.63$

d.f $\nu = n_1+n_2-2 = 9+7-2 = 14$

$t_{tab} = 1.761$

$t_c > t_{tab}$

H_0 is rejected

---

3. Two horses A and B are tested according to the time, to run a particular tract with the following results.

HORSE A	28	30	32	33	33	29	34
HORSE B	29	30	30	24	27	27	

Test whether you can discriminate between 2 horses.

NULL HYPOTHESIS: H_0:

There is no significant difference between 2 samples.

$\bar{X}_1 = 31.28$ $\bar{X}_2 = 27.8$

X_1	X_2	$(X_1-\bar{X}_1)^2$	$(X_2-\bar{X}_2)^2$
28	29	10.76	1.37
30	30	1.64	4.71
32	30	0.52	4.71
33	24	2.96	14.67
33	27	2.96	0.69
29	27	5.20	0.69
34		7.40	
		31.37	26.78

$$\Sigma_1 S^2 = \frac{31.44 + 26.84}{7+6-2} = \frac{58.28}{11} = 5.298$$

$$S = 2.3$$

$$t = \frac{31.28 - 27.83}{2.3\sqrt{\frac{1}{7}+\frac{1}{6}}} = \frac{3.45}{2.3 \times 0.56} = 2.67$$

$$|t|_{cal} = 2.67$$

$$d.f = \nu = n_1 + n_2 - 2 = 11$$

$$t_{tab} = 1.796$$

$$t_{cal} > t_{tab}$$

$\therefore$ H_0 is rejected.

Two horses can be discriminated.

F - test:

To test whether there is any significance between 2 estimates of population variance or to test if the 2 samples have come from the same population, we use F-test.

$$F = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

$$= \frac{S_1^2}{S_2^2}, \quad S_1^2 > S_2^2$$

$$d.f\ (\nu_1, \nu_2) = (n_1 - 1, n_2 - 1)$$

1. In one sample of 8 observation the sum of squares of deviations of sample value from the sample mean was 84.4, and in the other sample of 10 observation it was 102.6. Test whether this difference is significance at 5 percent level.

NULL HYPOTHESIS: H_0:

There is no significant difference at 5 percent level.

LoS = 5%

$$S_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{84.4}{8-1} = 12.057$$

$$S_2^2 = \frac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{102.6}{10-1} = 11.4$$

$$F_c = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} \simeq 1.06$$

2. The time taken by workers in performing a job by method I and method 2 are given bel[ow]

METHOD I	20	16	26	27	23	22	
METHOD II	27	33	42	35	32	34	38

Through the data show that the variances of time distribution from population, from which these samples are drawn do not differ significantly?

NULL HYPOTHESIS H_0 :

Variances of time distribution does not differ significantly.

$\overline{x}_1 = 22.3$ $\qquad$ $\overline{x}_2 = 34.4$

$(x_1 - \overline{x}_1)^2$	$(x_2 - \overline{x}_2)^2$
5.29	54.76
39.69	1.96
13.69	57.76
22.09	0.36
0.49	5.76
0.09	0.16
	12.96
81.34	133.72

$$S_1^2 = \frac{\Sigma(x_1-\overline{x}_1)^2}{n_1-1} = \frac{81.34}{6-1} = 16.26 \qquad S_2^2 = \frac{\Sigma(x_2-\overline{x}_2)^2}{n_2-1} = \frac{133.72}{7-1} = 22$$

$$F = \frac{\text{Greater variance}}{\text{Smaller variance}} = \frac{S_2^2}{S_1^2} = \frac{22.28}{16.26} = 1.37$$

$$d.f = (n_2-1, n_1-1) = (6, 5)$$

$$F_t = 4.95$$

$\therefore F_c < F_t$

$\therefore H_0$ is accepted

Variances of time distribution does not differ significantly.

3. Two random samples gave the following results,

SAMPLE	SIZE	SAMPLE MEAN	SUM OF SQUARE DEVIATION FROM THE MEAN
1.	10	15	90
2	12	14	108

Test whether the samples come from same normal population.

(i) $\mu_1 = \mu_2$ (ii) $\sigma_1^2 = \sigma_2^2$

(i) NULL HYPOTHESIS:

The samples come from same normal population.

i) $\mu_1 = \mu_2$ (ii) $\sigma_1^2 = \sigma_2^2$

(i) To test equality of mean : $\mu_1 = \mu_2$:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} , \quad S^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

$$S^2 = \frac{90 + 108}{10 + 12 - 2} = 9.9 , \quad S = 3.15$$

$$t = \frac{15 - 14}{3.15\sqrt{\frac{1}{10} + \frac{1}{12}}} = 0.74.$$

$d.f = n_1 + n_2 - 2 = 20$, $t_{tab} = 1.725$

$t_{cal} < t_{cal}$ ∴ H_0 is accepted. ∴ $\mu_1 = \mu_2$

(ii) To test equality of variances : $\sigma_1^2 = \sigma_2^2$

$$S_1^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{90}{10 - 1} = 10$$

$$S_2^2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{108}{12 - 1} = 9.81 , \quad F = \frac{S_1^2}{S_2^2} = \frac{10}{9.81} = 1.02$$

$d.f = (n_1 - 1, n_2 - 1) = (9, 11)$

$F_c < F_t$ ∴ H_0 is accepted.

∴ Samples comes from same normal population

CHI SQUARE TEST (χ^2)

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$O_i \rightarrow$ Observed values

$E_i \rightarrow$ Expected values

d.f $= n-1$

Note :

* In case of binomial distribution $\nu = n-1$
* In case of Poisson distribution $\nu = n-2$
* In case of Normal distribution $\nu = n-3$.

APPLICATIONS OF CHI-SQUARE TEST :

* Goodness of fit
* Independence of attribute.
* Test of homogenity.

1. The no. of automobile accidents per week in a certain community are as follows,

12, 8, 20, 2, 14, 10, 15, 6, 9, 4 .

Are these frequencies are in agreement with a belief that accident conditions were the same during this 10 week period.

NULL HYPOTHESIS H_0:

Accident conditions were the same during this 10 week period.

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
12	10	4	0.4
8	10	4	0.4
20	10	100	10
2	10	64	6.4
14	10	16	1.6
10	10	0	0
15	10	25	2.5
6	10	16	1.6
9	10	1	0.1
4	10	36	3.6
			$\Sigma = 26.6$

$$\chi^2_c = \Sigma\left[\frac{(O-E)^2}{E}\right] = 26.6$$

$$d.f = n-1 = 10-1 = 9$$

$$\chi^2_{tab} = 16.919$$

$$\chi^2_c > \chi^2_t$$

$\therefore$ H_0 is rejected.

$\therefore$ Accident conditions were not the same during this 10 week period.

2. The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

DIGITS	0	1	2	3	4	5	6	7	8	9
FREQUENCY	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory.

NULL HYPOTHESIS H_0 :

The digits occur equally frequently in the directory.

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1026	1000	676	0.676
1107	1000	11449	11.449
997	1000	9	0.009
966	1000	1156	1.156
1075	1000	5625	5.625
933	1000	4489	4.489
1107	1000	11449	11.449
972	1000	784	0.784
964	1000	1296	1.296
853	1000	21609	21.609
			58.542

d.f $= n-1 = 10-1 = 9$

$\chi^2_{tab} = 16.919$

$\chi^2_{cal} > \chi^2_{tab}$

$\therefore$ H_0 is rejected

3. The theory predicts that the proportion of beans in the four group A, B, C, D should be 9:3:3:1. In an experiment with 1600 beans, the numbers in the 4 groups were 882, 313, 287, 118.

Does the experimental result support the theory?

NULL HYPOTHESIS H_0:

The experimental result support the theory.

O	E	$(O-E)^2$	$(O-E)^2/E$
882	$\frac{3}{16} \times 1600 = 900$	324	0.36
313	$\frac{3}{16} \times 1600 = 300$	169	0.56
287	$\frac{3}{16} \times 1600 = 300$	169	0.56
118	$\frac{1}{16} \times 1600 = 100$	324	3.24
			4.72

$$\chi^2 = \sum \left[\frac{(O-E)^2}{E} \right] = 4.72$$

$d.f = n-1 = 4-1 = 3$

$\chi^2_{tab} = 7.815$

$\chi^2_{cal} < \chi^2_{tab}$

$\therefore H_0$ is accepted, i.e The experimental result support the theory.

---

4. A die is thrown 264 times with the following results. Show that the die is biased?

No. appreared on the dice	frequency
1	40
2	32
3	28
4	58
5	54
6	52

NULL HYPOTHESIS (H_0):

The die is unbiased

$$E = \frac{264}{6} = 44$$

O	E	$(O-E)^2$	$(O-E)^2/E$
40	44	16	0.36
32	44	144	3.27
28	44	256	5.82
58	44	196	4.45
54	44	100	2.27
52	44	64	1.45
			17.62

d.f $= n-1 = 6-1$

$= 5$

$\chi^2_{tab} = 11.070$

$\chi^2_{cal} > \chi^2_{tab}$

$\therefore$ H_0 is rejected

$\therefore$ The die is biased.

5. Fit a poisson distribution with the following data and test the goodness of fit

X	0	1	2	3	4	5	6
freq	275	72	30	7	5	2	1

NULL HYPOTHESIS(H_0):

The poisson distribution is a good fit to a given data

x	f	fx	$E = \frac{Ne^{-\lambda}\lambda^x}{x!}$
0	275	0	$392 \times e^{-0.48}(0.48)^0/0!$
1	72	72	116.43
2	30	60	27.9
3	7	21	4.47
4	5	20	0.536
5	2	10	0.05
6	1	6	0.0041
	392	189	

Any values <5, add till it becomes >5

$$\lambda = \bar{x} = \frac{\Sigma fx}{\Sigma f} = 0.48$$

O	E	$(O-E)^2/E$
275	242.56	4.34
72	116.43	16.95
30	27.9	0.16
15	5.06	19.52
		40.97

$$\chi^2 = \sum \frac{[O-E]^2}{E} = 40.95$$

d.f = n-2-3 (3 elements are added)

= 7-2-3 = 2

$\chi^2_{tab} = 5.991$

$\chi^2_{cal} > \chi^2_{tab}$

$\therefore H_0$ is rejected

$\therefore$ Poisson distribution is not a good fit to data.

χ^2 test for independence of attribute:

2×2 contigency table,

	B	
A	a	b
	c	d

Ho: the attributes are independent

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

degree of freedom (d.f) = $(r-1)(c-1)$

The expected frequencies are,

$\frac{(a+b)(a+c)}{N}$	$\frac{(a+b)(b+d)}{N}$	$a+b$
$\frac{(a+c)(c+d)}{N}$	$\frac{(b+d)(c+d)}{N}$	$c+d$
$a+c$	$b+d$	N

On the basis of information given below about the treatment of 200 patients suffering from a disease. State whether the new treatment is superior to conventional treatment.

	FAVOURABLE	NOT FAVOURABLE	TOTAL
NEW	60	30	90
CONVENTIONAL	40	70	110

H_0 : NULL HYPOTHESIS :

The new treatment and conventional treatment are independent.

$\frac{(60+30)(60+40)}{200} = 45$	$\frac{(60+30)(70+30)}{200} = 45$	$(a+b)$
$\frac{(60+40)(40+70)}{200} = 55$	$\frac{(60+40)(30+70)}{200} = 55$	$(c+d)$
$(a+c)$	$(b+d)$	$N = 200$

OBSVERVED VALUE(O)	E	$\frac{(O-E)^2}{E}$
60	45	5
30	45	5
40	55	4.09
70	55	4.09
		18.18

$\chi^2 cal = \sum_1 \left[\frac{(O-E)^2}{E}\right] = 18.18$

$d.f = (r-1)(c-1) = (2-1)(2-1) = 1$

$\chi^2 tab = 3.842$ (d.f = 1, los = 5%)

$\chi^2 cal > \chi^2 tab$

$\therefore H_0$ is rejected. The new treatment and conventional treatment are not independent.

2. The following table gives the classification of 100 workers based on gender and nature of work. Test whether the nature of work is independent of gender of the workers.

	STABLE	UNSTABLE	
MALES	40	20	60
FEMALES	10	30	40
	50	50	100

NULL HYPOTHESIS H_0:

The nature of work is independent of gender of the worker.

$\frac{60 \times 50}{100} = 30$	$\frac{60 \times 50}{100} = 30$	$(a+b)$
$\frac{50 \times 40}{100} = 20$	$\frac{40 \times 50}{100} = 20$	$(c+d)$
$(a+c)$	$(b+d)$	100

OBSERVED VALUE (O)	E	$\frac{(O-E)^2}{E}$
40	30	3.33
20	30	3.33
10	20	5
30	20	5
		16.66

$\chi^2_{cal} = 16.66$

$d.f = (r-1)(c-1) = (2-1)(2-1) = 1$

$\chi^2_{tab} = 3.841$ (5% level)

$\chi^2_{cal} > \chi^2_{tab}$

H_0 is rejected.

The nature of work is not independent of gender of the work.

3. Two researchers adopted different sampling techniques while investigating the same group of students to find number of students falling into different intelligent levels.

RESEARCHERS	BELOW AVERAGE	AVERAGE	ABOVE AVERAGE	GENIUS
X	86	60	44	10
Y	40	33	25	2

Would you say the sampling techniques adopted by two researchers are significantly different.

NULL HYPOTHESIS H_0:

There is no significant difference between the two reseachers.

EXPECTED FREQUENCIES:

RESEARCHERS	BA	A	AA	G	
X	84	62	46	8	200
Y	42	31	23	4	100
	126	93	69	12	300

O	E	$\frac{(O-E)^2}{E}$
86	84	0.048
60	62	0.065
44	46	0.087
10	8	0.5
40	42	0.095
33	31	0.129
27	27	0
		0.925

$$\chi^2_c = \sum \frac{(O-E)^2}{E} = 0.925$$

$$d.f = (c-1)(r-1) = (4-1)(2-1) - 1 = 2$$

χ^2_{tab} at 2 d.f $= 5.991$

$$\chi^2_c < \chi^2_t$$

$\therefore H_0$ is accepted.

In a certain sample of 2000 families 1400 families are consumers of tea out of 1800 Hindu families 1236 families consume tea. Use Chi square test and state whether there is any significant difference between consumption of tea among Hindu and non Hindu families.

OBSERVED FREQUENCY:

	HINDU	NON-HINDU	
FAMILIES CONSUMING TEA	1236	164	1400
FAMILIES NOT. CONS. TEA	564	36	600
	1800	200	2000

EXPECTED FREQUENCIES:

	HINDU	NON-HINDU
CONSUMING TEA	1260	140
NOT. CONS. TEA	540	60

O	E	$\frac{(O-E)^2}{E}$
1236	1260	0.457
164	140	4.114
564	540	1.067
36	60	9.6
		15.238

$d.f = (r-1)(c-1) = (2-1)(2-1) =$

$\chi^2_{tab} = 3.841$

$\chi^2_c > \chi^2_{tab}$

$\therefore H_0$ is rejected

LARGE SAMPLES:

TEST OF SIGNIFICANCE FOR SINGLE PROPORTION:

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{N}}} \quad , \; P + Q = 1$$

$p \rightarrow$ sample proportion

$P \rightarrow$ Population proportion.

1. In a sample of 1000 people in Karnataka, 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat eaters are equally popular in this state at 1% level?

$N = 1000$

NULL HYPOTHESIS H_0:

Rice eaters and wheat eaters are equally popular.

LOS: 1%

p = proportion of rice eaters out of $1000 = \frac{540}{1000}$

$P = \frac{1}{2}$, $Q = \frac{1}{2}$.

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{N}}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.53.$$

$|Z|_c = 2.53$

Z_t at $1\% = 2.58$

$|Z|_c < |Z|_t$

$\therefore H_0$ is accepted.

2. A manufacturer claims that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 182 where faulty. Test this claim at 5% level of significance.

NULL HYPOTHESIS H_0:

Atleast 95% of the equipment supplied to a factory conformed to specification.

LOS: 5%

$N = 200$. $p = \frac{182}{200} = 0.91$ $P = \frac{95}{100}$ $Q = \frac{5}{100}$

$$|Z|_c = \frac{p - P}{\sqrt{\frac{PQ}{N}}} = \frac{0.91 - 0.95}{\sqrt{\frac{(0.95 \times 0.05)}{200}}} = -2.59$$

$|z|_t = 1.96$

$Z_c > Z_t$

$\therefore$ H_0 is rejected

$\therefore$ The 95% of equipments supplied to a factory not conformed to specification.

3. A die was thrown 9000 times and of these 3220 yielded a three or four. Is this consistent with the hypothesis that the die is unbiased.

NULL HYPOTHESIS H_0:

Die is unbiased.

$N = 9000$

p = proportion of getting 3 or 4 $= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

P = Proportion of getting 3 or 4

$= \frac{3200}{9000} = 0.36$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{N}}} = \frac{0.36 - 0.33}{\sqrt{\frac{(0.33)(0.67)}{9000}}} = 6.05$$

$|z|_{cal} = 6.05$

$|z|_{tab} = 1.96$

$|z|_{cal} > |z|_{tab}$

$\therefore H_0$ is rejected.

$\therefore$ Die is biased.

4. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers.

NULL HYPOTHESIS : H_0:

There is no significant difference between smokers and non-smokers.

$N = 600$ $\quad P = \frac{1}{2} = 0.5$ $\quad Q = 0.5$ $\quad p = 0.542$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{N}}} = \frac{0.542 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = 1.96$$

$|z|_{cal} = 2.06$

$|z|_{tab} = 1.96$

$|z|_{cal} > |z|_{tab}$

$\therefore H_0$ is rejected.

5. A random sample of 500 pineapples were taken from a large consignment and 65 were found to be bad. Find the percentage of bad pineapples in the consignment

$n = 500$

p - proportion of bad pineapple

$p = \frac{65}{500} = 0.13$, $q = 0.87$

$$p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3\sqrt{\frac{(0.13)(0.87)}{500}}$$

$$= 0.085 \text{ to } 0.175$$

(i.e) 8.5% to 17.5%

II. TEST OF SIGNIFICANCE OF DIFFERENCE OF PROPORTION:

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \quad q = 1 - p$$

1. Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportion of men and women in favour of proposal are same.

NULL HYPOTHESIS H_0:

The proportion of men and women in favour of proposal are same.

$n_1 = 400$, $n_2 = 600$

Proportion of men $p_1 = \frac{200}{400} = \frac{1}{2} = 0.5$

Proportion of women $p_2 = \frac{325}{600} = 0.541$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400(0.5) + 600(0.541)}{1000} = 0.52$$

$q = 1 - p = 0.475$

Test statistics:

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.541}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}} = -1.3$$

$|z|_c = 1.3$

$|z|_{tab}$ at 5% $= 1.96$

$|z|_c < |z|_{tab}$

$\therefore H_0$ is accepted

The proportion of men and women in favour of proporsal are same.

2. In two large population there are 30% and 25% are fair haired people. Is this difference likely to be given in samples of 1200 and 900 respectively from the two population.

NULL HYPOTHESIS: H_0

The population proportions are equal.

$n_1 = 1200$ $\quad$ $n_2 = 900$

$P_1 = 0.30$ $\quad$ $P_2 = 0.25$

$Q_1 = 1 - P_1 = 0.7$ $\quad$ $Q_2 = 1 - P_2 = 0.75$

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

$$Z = \frac{0.3 - 0.25}{\sqrt{\frac{(0.3)(0.7)}{1200} + \frac{(0.75)(0.25)}{900}}} = 2.5$$

$|z|_c = 2.5$

$|z|_{tab}$ at $5\% = 1.96$

$|z|_c > |z|_t$

$\therefore H_0$ is rejected

The population proportions are not equal.

III. TEST OF SIGNIFICANCE FOR SINGLE MEAN:

$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

1. A sample of 900 members has a mean of 3.4 cm and standard deviation 2.61 cms. Is the sample from a large population of mean 3.25 cms and S.D 2.61 cms. If the population is normal and the mean is unknown, find the 95% fiducial limits.

NULL HYPOTHESIS: H_0:

The sample is form a large population of mean 3.25 cm and SD 2.61 cm.

L.O.S = 5%, $|z|_{tab} = 1.96$.

Test statistics :

$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}} = 1.724$$

$|z|_{cal} = 1.724$ $|z|_{tab} = 1.96$

$|z|_c < |z|_{tab}$

$\therefore H_0$ is accepted

The sample is from a large population of mean 3.25 cm and SD 2.61 cm

95% fiducial limits

$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}} = 3.4 \pm 1.96\left(\frac{2.61}{\sqrt{900}}\right) = (3.23, 3.57)$

2. An insurance agent has claimed that the average age of policy holders who issue through him is less that the average for all agents which 30.5 years. A random sample of 100 policy holders who had issued through him gave the following age distribution

AGE	16-20	21-25	26-30	31-35	36-40
NO. OF PERSONS	12	22	20	30	16

Calculate the Arithmetic mean and S.D of distribution and use these values to test this claim at 5% level of significance.

NULL HYPOTHESIS (H_0):

The average age of policy holders who issue through them is less than the average for all agents which is 30.5 years.

Level of significance = 5%.

$\mu = 30.5$

$i = 5$

X	m	f	$d = \frac{m-A}{i}$	fd	fd^2
15.5 – 20.5	18	12	-2	-24	48
20.5 – 25.5	23	22	-1	-22	22
25.5 – 30.5	28	20	0	0	0
30.5 – 35.5	33	30	1	30	30
35.5 – 40.5	38	16	2	32	64
				16	164

$$S = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2} \times 5 = 6.35$$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 28 + \frac{16}{100} \times 5 = 28 + 0.16 \times 5$$

Test statistics:

$$Z_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{28.8 - 30.5}{\frac{6.35}{\sqrt{100}}} = -2.67$$

$$|z|_{cal} = 2.67$$

Level of significance $= 5\%$

Z_{tab} at $5\% = 1.96$

$|z|_{cal} > |z|_{tab}$

$\therefore H_0$ is rejected.

The average age of policy holders who issue through them is not less than average for all agents is 30.5 years.

IV. TEST OF SIGNIFICANCE OF DIFFERENCE OF MEAN:

different S.D: $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

same S.D; If $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad \sigma = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

1. The means of two large samples with size 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can these samples be regarded as drawn from the same population of standard deviation 2.5 inches.

NULL HYPOTHESIS (H_0):

The samples are drawn from the same population of standard deviation 2.5 inches.

$n_1 = 1000$, $n_2 = 2000$, S.D $= 2.5$. $\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$

Test of statistics:

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}} = 5.16$$

$Z_{tab} = 1.96$ [los $= 5\%$]

$|z|_{cal} > |z|_{tab}$

$\therefore H_0$ is rejected.

The samples are not drawn from the same population of standard deviation 2.5 inches.

2. The mean yield of wheat from a district A was 210 pounds with SD 10 pounds per acre from a sample of 100 blocks. In another district B, the mean yield was 220 pounds with SD 12 pounds from a sample of 150 blocks. Assuming that the S.D of yield of entire stack was 11 pounds. Test whether there is any significance difference between mean yield of crops in two districts.

NULL HYPOTHESIS H_0:

There is no significance difference between mean yield of crops in two districts.

$n_1 = 100$ $\sigma_1 = 10$ $\bar{x}_1 = 210$

$n_2 = 150$ $\sigma_2 = 12$ $\bar{x}_2 = 220$

Test of Statistics:

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{210 - 220}{\sqrt{\frac{10^2}{100} + \frac{12^2}{150}}} = -7.14$$

$$|Z|_{cal} = 7.14$$

Level of significance = 5%

$$|Z|_{tab} = 1.96$$

$$|Z|_{cal} > |Z|_{tab}$$

$\therefore H_0$ is rejected.

There is a significance difference between mean yield of crops in two districts.

Chapter 415

Multivariate Analysis of Variance (MANOVA)

Introduction

Multivariate analysis of variance (MANOVA) is an extension of common analysis of variance (ANOVA). In ANOVA, differences among various group means on a single-response variable are studied. In MANOVA, the number of response variables is increased to two or more. The hypothesis concerns a comparison of vectors of group means. When only two groups are being compared, the results are identical to Hotelling's T^2 procedure.

The multivariate extension of the F-test is not completely direct. Instead, several test statistics are available, such as Wilks' Lambda and Lawley's trace. The actual distributions of these statistics are difficult to calculate, so we rely on approximations based on the F-distribution.

Technical Details

A MANOVA has one or more factors (each with two or more levels) and two or more dependent variables. The calculations are extensions of the general linear model approach used for ANOVA.

Unlike the univariate situation in which there is only one statistical test available (the F-ratio), the multivariate situation provides several alternative statistical tests. We will describe these tests in terms of two matrices, *H* and *E*. *H* is called the *hypothesis matrix* and *E* is the *error matrix*. These matrices may be computed using a number of methods. In **NCSS**, we use the standard general linear models (GLM) approach in which a sum of squares and cross-products matrix is computed. This matrix is based on the dependent variables and independent variables generated for each degree of freedom in the model. It may be partitioned according to the terms in the model.

MANOVA Test Statistics

For a particular p-variable multivariate test, assume that the matrices *H* and *E* have *h* and *e* degrees of freedom, respectively. Four tests may be defined as follows. See Seber (1984) for details. Let θ_i, ϕ_i, and λ_i be the eigenvalues of $H(E+H)^{-1}$, HE^{-1}, and $E(E+H)^{-1}$ respectively. Note that these eigenvalues are related as follows:

$$\theta_i = 1 - \lambda_i = \frac{\phi_i}{1 + \phi_i}$$

$$\phi_i = \frac{\theta_i}{1 - \theta_i} = \frac{1 - \lambda_i}{\lambda_i}$$

$$\lambda_i = 1 - \theta_i = \frac{1}{1 + \phi_i}$$

Wilks' Lambda

Define Wilks' Lambda as follows:

$$\Lambda_{p,h,e} = \frac{|E|}{|E + H|}$$

$$= \prod_{j=1}^{p} (1 - \theta_j)$$

with $e \geq p$.

The following approximation based on the F-distribution is used to determine significance levels:

$$F_{ph,ft-g} = \frac{(ft - g)\left(1 - \Lambda^{1/t}\right)}{ph\Lambda^{1/t}}$$

where

$$f = e - \frac{1}{2}(p - h + 1)$$

$$g = \frac{ph - 2}{2}$$

$$t = \begin{cases} \sqrt{\dfrac{p^2h^2 - 4}{p^2 + h^2 - 5}} & \text{if } p^2 + h^2 - 5 > 0 \\ 1 & \text{otherwise} \end{cases}$$

This approximation is exact if p or $h \geq 2$.

Lawley - Hotelling Trace

The trace statistic, T_g^2, is defined as follows:

$$T_g^2 = e\sum_{j=1}^{s}\phi_j$$

where

$$s = \min(p, h)$$

The following approximation based on the F-distribution is used to determine significance levels:

$$F_{a,b} = \frac{T_g^2}{ce}$$

where

$$a = ph$$

$$b = 4 + (a + 2)/(B - 1)$$

$$c = \frac{a(b - 2)}{b(e - p - 1)}$$

$$B = \frac{(e + h - p - 1)(e - 1)}{(e - p - 3)(e - p)}$$

Pillai's Trace

Pillai's trace statistic, $V^{(s)}$, is defined as follows:

$$V^{(s)} = \sum_{j=1}^{s}\theta_j = \mathrm{tr}(H(E + H)^{-1})$$

where

$$s = \min(p, h)$$

The following approximation based on the F-distribution is used to determine significance levels:

$$F_{s(2m+s+1),s(2n+s+1)} = \frac{(2n + s + 1)V^{(s)}}{(2m + s + 1)(s - V^{(s)})}$$

where

$$s = \min(p, h)$$

$$m = (|p - h| - 1)/2$$

$$n = (e - p - 1)/2$$

Roy's Largest Root

Roy's largest root, ϕ_{max}, is defined as the largest of the ϕ_i's. The following approximation based on the F-distribution is used to determine significance levels:

$$F_{(2_{v1}+2),(2_{v2}+2)} = \frac{2_{v2} + 2}{2_{v1} + 2}\phi_{max}$$

where

$$s = \min(p, h)$$

$$v_1 = (|p - h| - 1)/2$$

$$v_2 = (e - p - 1)/2$$

Which Test to Use

When the hypothesis degrees of freedom, *h*, is one, all four test statistics will lead to identical results. When *h*>1, the four statistics will usually lead to the same result. When they do not, the following guidelines from Tabachnick (1989) may be of some help.

Wilks' Lambda, Lawley's trace, and Roy's largest root are often more powerful than Pillai's trace if *h*>1 and one dimension accounts for most of the separation among groups. Pillai's trace is more robust to departures from assumptions than the other three.

Tabachnick (1989) provides the following checklist for conducting a MANOVA. We suggest that you consider these issues and guidelines carefully.

Assumptions and Limitations

The following assumptions are made when using a MANOVA.

1. The response variables are continuous.
2. The residuals follow the multivariate-normal probability distribution with means equal to zero.
3. The variance-covariance matrices of each group of residuals are equal.
4. The individuals are independent.

Multivariate Normality and Outliers

MANOVA is robust to modest amount of skewness in the data. A sample size that produces 20 degrees of freedom in the univariate F-test is adequate to ensure robustness. Non-normality caused by the presence of outliers can cause severe problems that even the robustness of the test will not overcome. You should screen your data for outliers and run it through various univariate and multivariate normality tests and plots to determine if the normality assumption is reasonable.

Homogeneity of Covariance Matrices

MANOVA makes the assumption that the within-cell (group) covariance matrices are equal. If the design is balanced so that there is an equal number of observations in each cell, the robustness of the MANOVA tests is guaranteed. If the design is unbalanced, you should test the equality of covariance matrices using Box's M test. If this test is significant at less than 0.001, there may be severe distortion in the alpha levels of the tests. You should only use Pillai's trace criterion in this situation.

Linearity

MANOVA assumes linear relationships among the dependent variables within a particular cell. You should study scatter plots of each pair of dependent variables using a different color for each level of a factor. Look carefully for curvilinear patterns and for outliers. The occurrence of curvilinear relationships will reduce the power of the MANOVA tests.

Multicollinearity and Singularity

Multicollinearity occurs when one dependent variable is almost a weighted average of the others. This collinearity may only show up when the data are considered one cell at a time. The R^2-Other Y's in the Within-Cell Correlations Analysis report lets you determine if multicollinearity is a problem. If this R^2 value is greater than .99 for any variable, you should take corrective action (remove one of the variables). To correct for multicollinearity, begin removing the variables one at a time until all of the R^2's are less than .99. Do not remove them all at once! *Singularity* is the extreme form of multicollinearity in which the R^2 value is one.

Forms of multicollinearity may show up when you have very small cell sample sizes (when the number of observations is less than the number of variables). In this case, you must reduce the number of dependent variables.

Data Structure

The data must be entered in a format that places the dependent variables and values of each factor side by side. An example of the data for a MANOVA design is shown in the table below. In this example, *WRATR* and *WRATA* are the two dependent variables. *Treatment* and *Disability* are two factor variables. This database is stored in the file MANOVA1.

MANOVA1 Dataset (Subset)

WRATR	WRATA	Treatment	Disability
115	108	1	1
98	105	1	1
107	98	1	1
90	92	2	1
85	95	2	1
80	81	2	1
100	105	1	2
105	95	1	2
95	98	1	2
70	80	2	2
.	.	.	.
.	.	.	.
.	.	.	.

Unequal Sample Size and Missing Data

You should begin by screening your data. Pay particular attention to patterns of missing values. When using MANOVA, you should have more observations per factor category than you have dependent variables so that you can test the equality of covariance matrices using Box's M test.

NCSS ignores rows with missing values. If it appears that most of the missing values occur in one or two variables, you might want to leave these out of the analysis in order to obtain more data and hence more power.

NCSS uses the GLM procedure for calculating the hypothesis and error matrices. Each matrix is calculated as if it were fit last in the model. This is the recommended way of obtaining these matrices. This method is valid even when the sample sizes for the various groups are unequal.

Example 1 – Multivariate Analysis of Variance

This section presents an example of how to run an analysis of the data contained in the MANOVA1 dataset.

Setup

To run this example, complete the following steps:

1 Open the MANOVA1 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **MANOVA1** and click **OK**.

2 Specify the Multivariate Analysis of Variance (MANOVA) procedure options

- Find and open the **Multivariate Analysis of Variance (MANOVA)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variables....................................... **WRATR-WRATA**
Factor Variable 1 .... **Treatment**
Factor Variable 2 .... **Disability**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Expected Mean Squares

Expected Mean Squares

Source Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A: Treatment	1	Yes	S(AB)	S+bsA
B: Disability	2	Yes	S(AB)	S+asB
AB	2	Yes	S(AB)	S+sAB
S(AB)	12	No		S

Note: Expected Mean Squares are for the balanced cell-frequency case.

The Expected Mean Square expressions are provided to show the appropriate error term for each factor. The correct error term for a factor is that term that is identical except for the factor being tested.

Source Term

The source of variation or term in the model.

DF

The degrees of freedom. The number of observations "used" by this term.

Term Fixed?

Indicates whether the term is "fixed" or "random."

Denominator Term

Indicates the term used as the denominator in the F-ratio.

Expected Mean Square

This expression represents the expected value of the corresponding mean square if the design was completely balanced. "S" represents the expected value of the mean square error (sigma). The uppercase letters represent either the adjusted sum of squared treatment means if the factor is fixed, or the variance component if the factor is random. The lowercase letter represents the number of levels for that factor, and "s" represents the number of replications of the experimental layout.

These EMS expressions are provided to determine the appropriate error term for each factor. The correct error term for a factor is that term whose EMS is identical except for the factor being tested.

MANOVA Tests

Term(DF): Name Test Statistic	Test Value	DF1	DF2	F-Ratio	P-Value	Decision (α = 0.05)
A(1): Treatment						
Wilks' Lambda	0.137721	2	11	34.44	0.000018	Reject
Lawley - Hotelling Trace	6.261036	2	11	34.44	0.000018	Reject
Pillai's Trace	0.862279	2	11	34.44	0.000018	Reject
Roy's Largest Root	6.261036	2	11	34.44	0.000018	Reject
F (WRATR)	2090.888889	1	12	46.12	0.000019	Reject
F (WRATA)	1494.222222	1	12	33.25	0.000089	Reject
B(2): Disability						
Wilks' Lambda	0.255263	4	22	5.39	0.003528	Reject
Lawley - Hotelling Trace	2.895034	4	20	7.24	0.000896	Reject
Pillai's Trace	0.750481	4	24	3.60	0.019460	Reject
Roy's Largest Root	2.887241	2	12	17.32	0.000290	Reject
F (WRATR)	260.388889	2	12	5.74	0.017784	Reject
F (WRATA)	563.388889	2	12	12.54	0.001151	Reject
AB(2)						
Wilks' Lambda	0.908068	4	22	0.27	0.893037	Accept
Lawley - Hotelling Trace	0.100954	4	20	0.25	0.904790	Accept
Pillai's Trace	0.092192	4	24	0.29	0.881598	Accept
Roy's Largest Root	0.098039	2	12	0.59	0.570550	Accept
F (WRATR)	1.055556	2	12	0.02	0.977029	Accept
F (WRATA)	26.388889	2	12	0.59	0.571116	Accept

This report gives the results of the various significance tests. Usually, the four multivariate tests will lead to the same conclusions. When they do not, refer to the discussion of these tests found earlier in this chapter.

Once a multivariate test has found a term significant, use the univariate ANOVA to determine which of the variables and factors are “causing” the significance.

Term(DF): Name

The term in the design model with the degrees of freedom of the term in parentheses. For main effects, the variable name is also given.

Test Statistic

The name of the statistical test shown on this row of the report. The four multivariate tests are followed by the univariate F-tests of each variable.

Test Value

The value of the test statistic.

DF1

The numerator degrees of freedom of the F-ratio corresponding to this test.

DF2

The denominator degrees of freedom of the F-ratio corresponding to this test.

F-Ratio

The value of the F-test corresponding to this test. In some cases, this is an exact test. In other cases, this is an approximation to the exact test. See the discussion of each test to determine if it is exact or approximate.

P-Value

The significance level of the above F-ratio. The probability of an F-ratio larger than that obtained by this analysis. For example, to test at an alpha of 0.05, this probability would have to be less than 0.05 to make the F-ratio significant.

Decision (α = 0.05)

The decision to accept or reject the null hypothesis at the given level of significance. Note that you specify the level of significance when you select Alpha.

Correlations and Within-Group Covariances

Correlations and Within-Group Covariances

	WRATR	WRATA
WRATR	45.33333	2.583333
WRATA	0.0572313	44.94444

The correlations are shown in the off-diagonals in the lower-left half of the matrix.
The within-group covariances are shown on the diagonal and in the off-diagonals in the upper-right half of the matrix.

This report displays the correlations and covariances formed by averaging across all of the individual group covariance matrices. The correlations are shown in the lower-left half of the matrix. The within-group covariances are shown on the diagonal and in the upper-right half of the matrix.

Within-Cell Correlations Analysis

Within-Cell Correlations Analysis

Variable	R-Squared Other Y's	Canonical Variate	Eigenvalue	Percent of Total	Cumulative Total
WRATR	0.003275	1	1.057231	52.86	52.86
WRATA	0.003275	2	0.942769	47.14	100.00

This report analyzes the within-cell correlation matrix. It lets you diagnose multicollinearity problems as well as determine the number of dimensions that are being used. This is useful in determining if Pillai's trace should be used.

R-Squared Other Y's

This is the R-Squared index of this variable with the other variables. When this value is larger than 0.99, severe multicollinearity problems exist. If this happens, you should remove the variable with the largest R-Squared and re-run your analysis.

Canonical Variate

The identification numbers of the canonical variates that are generated during the analysis. The total number of variates is the smaller of the number of variables and the number of degrees of freedom in the model.

Eigenvalue

The eigenvalues of the within correlation matrix. Note that this value is not associated with the variable at the beginning of the row, but rather with the canonical variate number directly to the left.

Percent of Total

The percent that the eigenvalue is of the total. Note that the sum of the eigenvalues will equal the number of variates. If the percentage accounted for by the first eigenvalue is relatively large (70 or 80 percent), Pillai's trace will be less powerful than the other three multivariate tests.

Cumulative Total

The cumulative total of the Percent of Total column.

Univariate Analysis of Variance (ANOVA) Tables

Analysis of Variance (ANOVA) Table for WRATR

Source Term	DF	Sum of Squares	Mean Mean Square	F-Ratio	P-Value	Power (α = 0.05)
A: Treatment	1	2090.889	2090.889	46.12	0.000019*	0.999988
B: Disability	2	520.7778	260.3889	5.74	0.017784*	0.763859
AB	2	2.111111	1.055556	0.02	0.977029	0.052757
S	12	544	45.33333			
Total (Adjusted)	17	3157.778				
Total	18					

* Term significant at alpha = 0.05

Analysis of Variance (ANOVA) Table for WRATA

Source Term	DF	Sum of Squares	Mean Mean Square	F-Ratio	P-Value	Power (α = 0.05)
A: Treatment	1	1494.222	1494.222	33.25	0.000089*	0.999519
B: Disability	2	1126.778	563.3889	12.54	0.001151*	0.981144
AB	2	52.77778	26.38889	0.59	0.571116	0.125682
S	12	539.3333	44.94444			
Total (Adjusted)	17	3213.111				
Total	18					

* Term significant at alpha = 0.05

This is the standard ANOVA report as documented in the General Linear Models chapter. A separate report is displayed for each of the response variables.

Means and Standard Errors

Means and Standard Errors of WRATR

Term	Count	Mean	Standard Error*
All	18	89.11111	
A: Treatment			
1	9	99.88889	2.234687
2	9	78.33334	2.234687
B: Disability			
1	6	95.83334	2.736922
2	6	88.83334	2.736922
3	6	82.66666	2.736922
AB: Treatment, Disability			
1, 1	3	106.6667	3.870592
1, 2	3	100	3.870592
1, 3	3	93	3.870592
2, 1	3	85	3.870592
2, 2	3	77.66666	3.870592
2, 3	3	72.33334	3.870592

* The standard errors are calculated using the mean square error from the corresponding Analysis of Variance (ANOVA) table.

Means and Standard Errors of WRATA

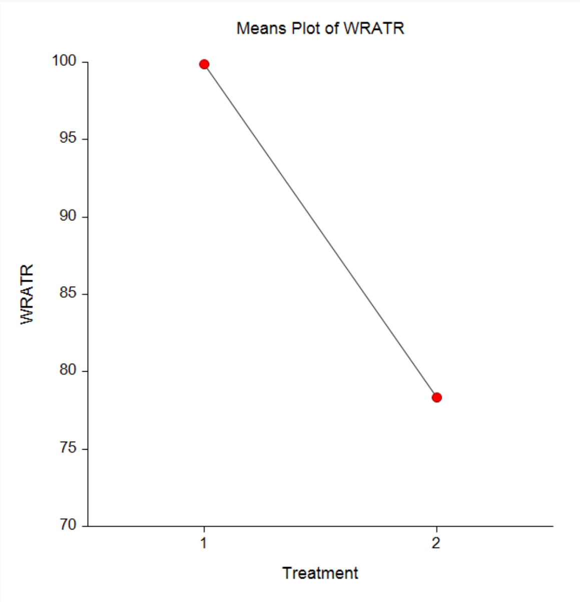
Term	Count	Mean	Standard Error*
All	18	87.22222	
A: Treatment			
1	9	96.33334	2.234687
2	9	78.11111	2.234687
B: Disability			
1	6	96.5	2.736922
2	6	88	2.736922
3	6	77.16666	2.736922
AB: Treatment, Disability			
1, 1	3	103.6667	3.870592
1, 2	3	99.33334	3.870592
1, 3	3	86	3.870592
2, 1	3	89.33334	3.870592
2, 2	3	76.66666	3.870592
2, 3	3	68.33334	3.870592

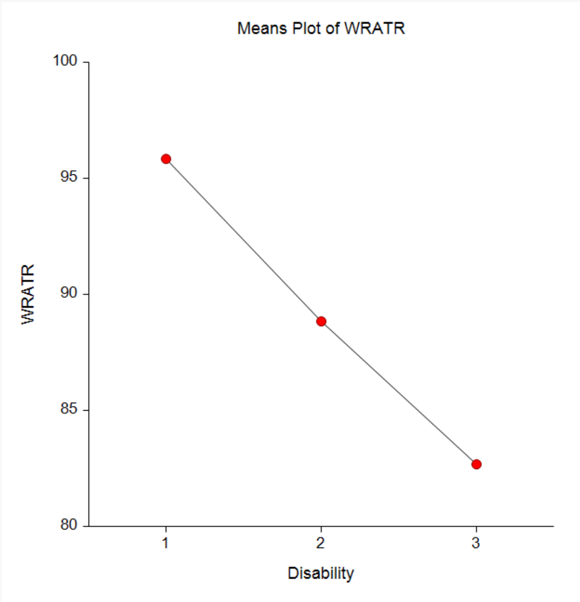
* The standard errors are calculated using the mean square error from the corresponding Analysis of Variance (ANOVA) table.

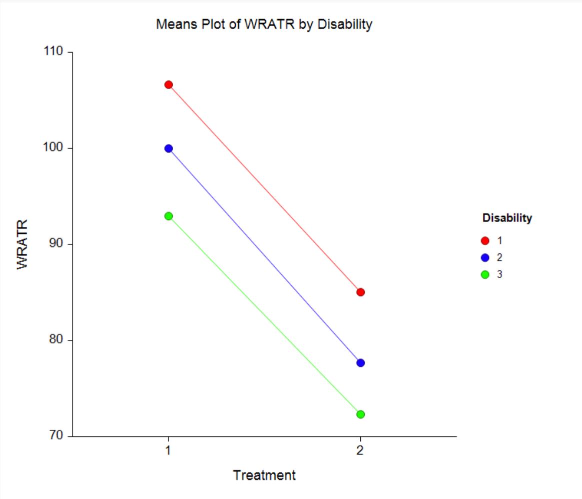
This report provides the least-squares means and standard errors for each variable. Note that the standard errors are calculated using the mean square error from the ANOVA table. They are not the standard errors that would be calculated from the individual cells.

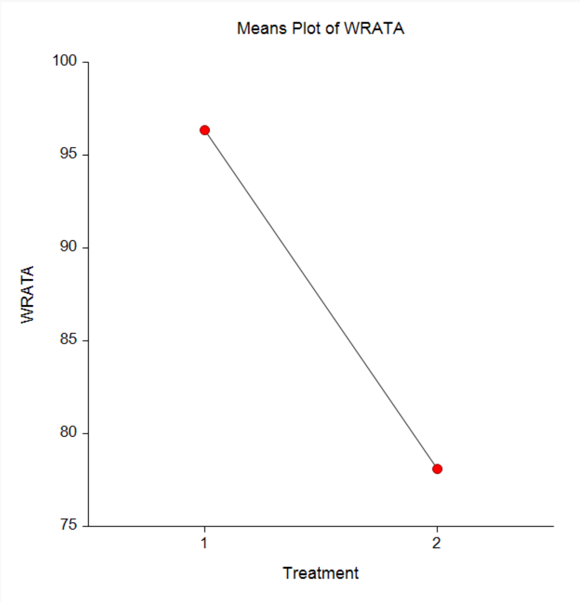
Means Plots

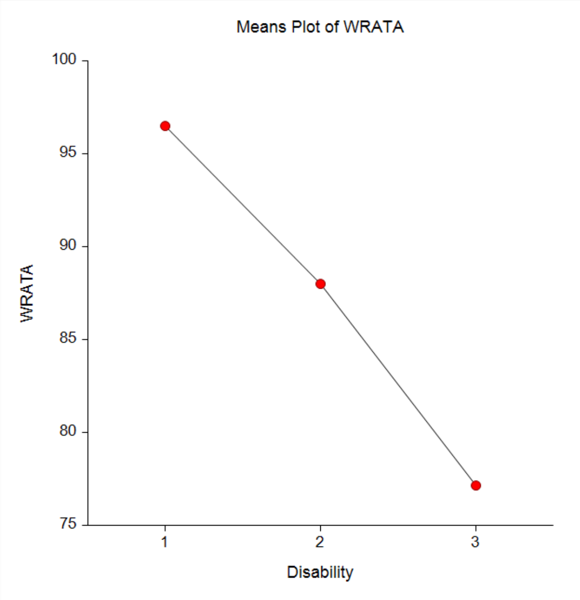
Means Plots

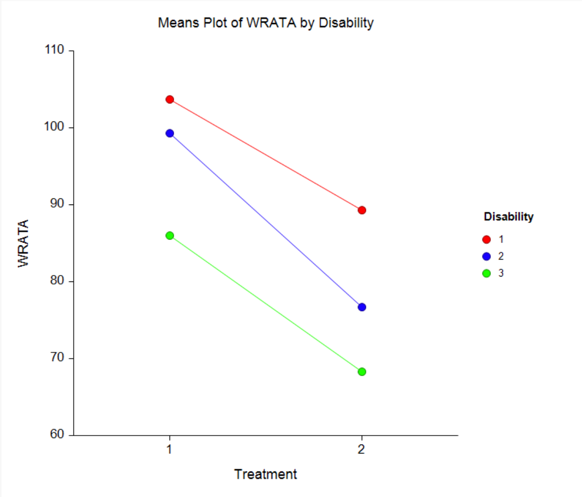












This section provides the means plots for main effects and interactions.