Unit 1
Statistical Techniques I

MCQ (1 Marks)

1. The first three raw moments of a distribution are 0, 2.5, 0.7. Find the value of the second central moment: (CO1)
   A. 0
   B. 1
   C. 2.5
   D. 0.7
2. The first three central moments of a distribution are 0, 15, -31. Find the moment coefficient of skewness. (CO1)
   A. 0.2847
   B. 2.847
   C. 28.47
   D. 284.7
3. The mean of 200 items was 50. Later on it was discovered that two items were misread as 92 and 8 instead of 192 and 88. Find out the correct mean. (CO1)
   A. 50.2
   B. 50.7
   C. 50.4
   D. 50.9
4. The spearman rank correlation coefficient is given by-(CO1)
   A. $r = 1 - 6\frac{\Sigma d^2}{n(n^2-1)}$
   B. $r = 1 - 6\frac{\Sigma d^2}{(n^2-n)}$
   C. $r = 1 - \frac{\Sigma d^2}{n(n^2-1)}$
   D. $r = 1 - 6\frac{\Sigma d^2}{(n^3-1)}$
5. If the regression coefficients are 0.8 and 0.2, what would be the value of coefficient of correlation? (CO1)
   A. 0.14
   B. 0.8
   C. 0.4
   D. 0.02
6. Curve which are more sharply peaked that normal curve is called: (CO1)
   A. Mesokurtic Curve
   B. Leptokurtic Curve
   C. Platykurtic Curve
   D. None of these.

7. Two lines of regression are $x + 2y - 5 = 0, 2x + 3y - 8 = 0$ then mean value of x and y are respectively: (CO1)
   A. 4,7
   B. 1,2
   C. -1, -2
   D. None of these.

Very Short Answer Type Questions (2 Marks)

8. Karl Pearson's coefficient of skewness of a distribution is 0.32, its standard deviation is 6.5 and mean is 29.6. Find the mode of the distribution. (CO1)

Solution: $SKP = \frac{A.M. - Mode}{SD} \Rightarrow 0.32 = \frac{29.6 - Mode}{6.5}$

$29.6 - Mode = (0.32 \times 6.5) = 2.08$

$Mode = 29.6 - 2.08 = 27.52$ Ans.

9. Write the angle between two lines of regression. (CO1)

Solution: $\tan\theta = \frac{1 - r^2}{r}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$

10. Find the Arithmetic mean of the following data. (CO1)

Class interval	0-10	10-20	20-30	30-40	40-50
Frequency	10	20	40	20	10

Solution:

C.I.	f_i	x_i	$f_i x_i$
0-10	10	5	50
10-20	20	15	300
20-30	40	25	1000
30-40	20	35	700
40-50	10	45	450
	100		2500

$A.M. = \frac{\Sigma f_i x_i}{\Sigma f_i}$

$= \frac{2500}{100}$

$= 25$ Ans.

11. Write the normal equation of $y = a + \frac{b}{x}$ (CO1)

Solution: $y = a + \frac{b}{x}$

$\Sigma y = na + b\Sigma x$

$\Sigma \frac{y}{x} = \frac{a}{\Sigma x} + nb$ Ans.

Short Answer Type Questions (6 Marks)

12. The first four moments about the working mean 28.5 of a distribution are 0.294, 7.144, 42.409 and 454.98. Calculate the moments about mean and measures of the skewness and kurtosis of the distribution. (CO1)

Solution: $A = 28.5;\ \mu_1' = 0.294;\ \mu_2' = 7.144;\ \mu_3' = 42.409$

$\mu_1 = 0;\ \mu_2 = \mu_2' - \mu_1'^2 = 7.0576$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 36.1588$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = 408.7895$

① $S_m = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{36.1588}{\sqrt{(7.0576)^3}} = 1.92970 \Rightarrow$ skewed

② $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{408.7895}{(7.0576)^2} = 8.2084737 3 \Rightarrow$ Leptokurtic

13. If the coefficient of correlation between two variables x and y is 0.5 and the acute angle between their lines of regression is $\tan^{-1}(3/5)$. show that $\sigma_x = \sigma_y/2$.(CO1) $r = 0.5;\ \theta = \tan^{-1}\left(\frac{3}{5}\right) \Rightarrow \tan\theta = \frac{3}{5}$

Solution: $\tan\theta = \frac{1-r^2}{r}\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right) \Rightarrow \frac{3}{5} = \frac{\left(1-\frac{1}{4}\right)}{\left(\frac{1}{2}\right)}\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$

$\Rightarrow \frac{3}{5} = \frac{3}{2}\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right) \Rightarrow 2\sigma_x^2 + 2\sigma_y^2 - 5\sigma_x\sigma_y = 0.$

$2\sigma_x^2 - 4\sigma_x\sigma_y - \sigma_x\sigma_y + 2\sigma_y^2 = 0.$

$(2\sigma_x - \sigma_y)(\sigma_x - 2\sigma_y) = 0.$

$\sigma_x - 2\sigma_y = 0 \Rightarrow \sigma_x = 2\sigma_y$

$2\sigma_x - \sigma_y = 0 \Rightarrow \boxed{\sigma_x = \frac{\sigma_y}{2}}$ Hence, Proved.

14. By the method of least squares, fit the curve $y = ax+bx^2$ to the following data: (CO1)

x	1	2	3	4	5
y	1.8	5.1	8.9	14.1	19.8

Solution:

$y = ax + bx^2$

$\Sigma yx = a\Sigma x^2 + b\Sigma x^3$

$\Sigma yx^2 = a\Sigma x^3 + b\Sigma x^4$

x	y	xy	x^2	x^3	x^2y	x^4
1	1.8	1.8	1	1	1.8	1
2	5.1	10.2	4	8	20.4	16
3	8.9	26.7	9	27	80.1	81
4	14.1	56.4	16	64	225.6	256
5	19.8	99.0	25	125	495	625
15	49.7	194.1	55	225	822.9	979

$194.1 = 55a + 225b$ — (1)

$822.9 = 225a + 979b$

$\boxed{a = 1.512}$

$\boxed{b = 0.49}$

$\boxed{y = (1.512)x + (0.49)x^2}$

15. For two random variables, x and y with same mean, the two regression equations are $y = ax + b$ and $x = \alpha y + \beta$. Show that $\frac{b}{\beta} = \frac{1-a}{1-\alpha}$. Also find Common mean. (CO1) $\boxed{\bar{x} = \bar{y} = m}$

Solution:

$y = ax + b$ — (3)

$\therefore b_{yx} = a$

i.e., 'y on x':

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

$y - m = a(x - m)$

$y = ax + m(1-a)$ — (1)

Now; $x = \alpha y + \beta$ — (4)

$b_{xy} = \alpha$

$\therefore$ 'x on y':

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

$x - m = \alpha(y - m)$

$x = \alpha y + m(1-\alpha)$ — (2)

Comparing (1) and (3);

$b = m(1-a)$

Comparing (2) and (4);

$\beta = m(1-\alpha)$

$\boxed{\frac{b}{\beta} = \frac{1-a}{1-\alpha}}$ Hence, Proved.

16. Find the coefficient of correlation for the following data: (CO1)

X	10	14	18	22	26	30
Y	18	12	24	6	30	36

Solution:

Let $u = \frac{x-a}{h}$; $v = \frac{y-b}{k}$

$\boxed{r_{uv} = r_{xy}}$

x	y	$(x-22)$	$u = \frac{x-22}{4}$	$(y-24)$	$v = \frac{y-24}{6}$	uv
10	18	−12	−3	−6	−1	3
14	12	−8	−2	−12	−2	4
18	24	−4	−1	0	0	0
22	6	0	0	−18	−3	0
26	30	4	1	6	1	1
30	36	8	2	12	2	4
			−3		−3	12

$\Sigma u^2 = 19$; $\Sigma v^2 = 19$

$r_{uv} = \frac{6(12) - (-3)(-3)}{\sqrt{6(19)-9}\,\sqrt{6(19)-9}}$

$\boxed{r_{uv} = 0.6}$ Ans.

17. The following table gives age(x) in years of cars and annual maitenance(y) in hundered rupees: (CO1)

x	1	3	5	7	9
y	15	18	21	23	22

Estimate the maintenance cost for a 4 year old car after finding the regression equation.

Solution:

x	y	xy	x^2
1	15	15	1
3	18	54	9
5	21	105	25
7	23	161	49
9	22	198	81
25	99	533	165

'y on x':

$$b_{yx} = \frac{5(533)-(25)(99)}{5(165)-(25)^2}$$

$$= \frac{190}{200} = 0.95$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{25}{5} = 5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{99}{5} = 19.8$$

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

$$\Rightarrow (y-\bar{y}) = 0.95(x-\bar{x})$$

$$y-19.8 = 0.95(x-5)$$

when $x = 4$;

$$y-19.8 = 0.95(-1)$$

$$\Rightarrow y = -0.95+19.8$$

$$\boxed{y = 18.85} \text{ Ans.}$$

18. The first four moments of a distribution about 2 are 1,2.5,5.5 and 16 respectively. Calculate the four moments about mean and about the origin. (CO1) $A=2$; $\mu_1'=1$; $\mu_2'=2.5$; $\mu_3'=5.5$; $\mu_4'=16$

Solution: about mean:

$$\mu_1' = \bar{x} - A$$

$$1 = \bar{x}-2$$

$$\bar{x} = 3$$

$$\mu_1 = 0 \; ; \; \mu_2 = \mu_2' - \mu_1'^2 = 2.5-1 = 1.5$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 5.5-3(2.5)(1)+2 = 0$$

$$\mu_4 = 16-4(5.5)(1)+6(2.5)(1)^2-3(1)^4 = 6$$

about origin:

$$\nu_1 = \bar{x} = 3 \; ; \; \nu_2 = \mu_2 + \bar{x}^2 = 1.5+9 = 10.5$$

$$\nu_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3 = 0+3(1.5)(3)+3^3 = 40.5$$

$$\nu_4 = \mu_4 + 3\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4$$

$$= 6+0+6(1.5)(9)+3^4 = 168$$

Long Answer Type Questions (10 Marks)

19. Find the moment coeff. of Skewness and kurtosis of the following data. (CO1)

Class- interval	0-10	10-20	20-30	30-40	40-50
Frequency	10	20	40	20	10

Solution:

CI	f_i	x_i	f_ix_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$(x_i-\bar{x})^4$	$f_i(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^3$	$f_i(x_i$
0-10	10	5	50	-20	400	8000	160000	4000	80000	1600
10-20	20	15	300	-10	100	1000	10000	2000	20000	2000
20-30	40	25	1000	0	0	0	0	0	0	0
30-40	20	35	700	10	100	10000	10000	2000	20000	2000
40-50	10	45	450	20	400	8000	160000	4000	80000	1600
	100		2500					12000	200000	3600

$$M = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{2500}{100} = 25 \; ; \; \mu_1 = 0.$$

$$\mu_2 = \frac{\Sigma f_i (x_i-\bar{x})^2}{\Sigma f_i} = \frac{12000}{100} = 120$$

$$\mu_3 = \frac{\Sigma f_i (x_i-\bar{x})^3}{\Sigma f_i} = \frac{200000}{100} = 2000$$

$$\mu_4 = \frac{\Sigma f_i (x_i-\bar{x})^4}{\Sigma f_i} = \frac{3600000}{100} = 36000$$

(1) $S_{KM} = \dfrac{\mu_3}{\sqrt{\mu_2^3}} = \dfrac{2000}{\sqrt{(120)^3}} = \dfrac{2000}{1314.534} = 1.521470$

+vely skewed.

(2) $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{36000}{120 \times 120} = 2.5 < 3$

Platykurtic.

20. The pressure of the gas corresponding to various volumes V is measures given by the following data

V(cm³)	50	60	70	90	100
P(kg cm⁻²)	64.7	51.3	40.5	25.9	78

Fit the data to equation $PV^{\gamma} = C$. (CO1)

Solution: $PV^{\gamma} = c \Rightarrow \log P + \gamma \log V = \log c$

$\underline{\log P} = \underline{\log c} - \gamma \log V$

$Y = A + BX$, $B \to -\gamma$

$\Sigma y = nA + B\Sigma x$

$\Sigma xy = A\Sigma x + B\Sigma x^2$

V	P	X = log V	Y = log P	XY	X^2
50	64.7	1.6989	1.8109	3.076	2.8862
60	51.3	1.7781	1.7101	3.0407	3.1616
70	40.5	1.8450	1.6074	2.965	3.4040
90	25.9	1.9542	1.4132	2.761	3.8188
100	78	2	1.8920	3.784	4
		9.2762	8.4336	15.626	17.2706

$8.4336 = 5A + 9.2762B$ — (1)

$15.626 = 9.2762A + 17.2706B$ — (2)

$\boxed{A = 2.305}$

$\boxed{B = -0.3335}$

$Y = 2.305 - 0.3335X$

$\log P = 2.305 - 0.3335 \log V$ Ans.

21. Use the method of Least squares to fit the curve: $y = \frac{c_0}{x} + c_1\sqrt{x}$ to the following data. (CO1)

x	0.1	0.2	0.4	0.5	1	2
y	21	11	7	6	5	6

Solution:

$$y = \frac{c_0}{x} + c_1\sqrt{x}$$

$$\Sigma \frac{y}{x} = c_0 \Sigma \frac{1}{x^2} + c_1 \Sigma \frac{1}{\sqrt{x}}$$

$$\Sigma y\sqrt{x} = c_0 \Sigma \frac{1}{\sqrt{x}} + c_1 \Sigma x$$

x	y	$\frac{y}{x}$	$\frac{1}{x^2}$	$\frac{1}{\sqrt{x}}$	$y\sqrt{x}$
0.1	21	210	100	3.16	6.64
0.2	11	55	25	2.23	4.91
0.4	7	17.5	6.25	1.58	4.42
0.5	6	12	4	1.41	4.24
1	5	5	1	1	5
2	6	3	0.25	0.70	8.48
4.2	56	302.5	136.5	10.08	33.69

$$302.5 = c_0(136.5) + 10.08\,c_1$$

$$33.69 = 10.08\,c_0 + 4.2\,c_1$$

$$\boxed{c_0 = 1.97} \qquad \boxed{c_1 = 3.28}$$

$$y = \frac{c_0}{x} + c_1\sqrt{x} = \frac{1.97}{x} + 3.28\sqrt{x}$$ Ans.

22. In a partially destroyed laboratory record of analysis of a correlation data, the following results only are legible: (CO1)
Variance of $x = 9$.
Regression equations: $8x - 10y + 66 = 0, 40x - 18y = 214$.
What were (a) the mean values of x and y (b) the standard deviation of y and the coefficient correlation between x and y.

Solution: Regression lines intersect at $(\bar{x}, \bar{y})$. So,

$8\bar{x} - 10\bar{y} + 66 = 0$; $40\bar{x} - 18\bar{y} = 214$

(a) $\boxed{\text{Mean is } (13, 17) \Rightarrow \bar{x} = 13, \bar{y} = 17}$

(b) $8x - 10y + 66 = 0$

Let x on y:

$8x = 10y - 66$

$x = \frac{10}{8}y - \frac{66}{8}$

$\boxed{b_{xy} = \frac{10}{8}}$

$\Rightarrow 40x - 18y - 214 = 0$

Let y on x:

$18y = 40x - 214$

$y = \frac{40}{18}x - \frac{214}{18}$

$\boxed{b_{yx} = \frac{40}{18}}$

$b_{yx} \cdot b_{xy} = r^2$

$\frac{10}{8} \times \frac{40}{18} = r^2$

$\boxed{r = 1.667}$ X

$\because -1 \leq r \leq 1$

$\therefore$ it is not correct

$8x - 10y + 66 = 0$

Let y on x:

$10y = 8x + 66$

$y = \frac{8x}{10} + \frac{66}{10}$

$\boxed{b_{yx} = \frac{8}{10}}$

$\Rightarrow 40x - 18y - 214 = 0$

Let x on y:

$40x = 18y + 214$

$x = \frac{18y}{40} + \frac{214}{40}$

$\boxed{b_{xy} = \frac{18}{40}}$

$b_{xy} \times b_{yx} = r^2$

$\boxed{r = 0.6}$

$-1 \leq r \leq 1$

$\therefore$ it is right

$\sigma_x^2 = 9$

$\sigma_x = 3$

$b_{yx} = \frac{r\sigma_y}{\sigma_x}$

$\frac{8}{10} = \frac{0.6(\sigma_y)}{3}$

$\boxed{\sigma_y = 4}$

So, standard deviation of y is 4.

23. Find the multiple linear regressions of x on y and z from the data relating to three variables: (CO1)

x	4	6	7	9	13	15
y	15	12	8	6	4	3
z	30	24	20	14	10	4

Solution:

x	y	z	xy	y^2	yz	xz	z^2
4	15	30	60	225	450	120	900
6	12	24	72	144	288	144	576
7	8	20	56	64	160	140	400
9	6	14	54	36	84	126	196
13	4	10	52	16	40	130	100
15	3	4	45	9	12	60	16
54	48	102	339	494	1034	720	2188

$$x = a + by + cz$$

$$\Sigma x = na + b\Sigma y + c\Sigma z$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2 + c\Sigma yz$$

$$\Sigma xz = a\Sigma z + b\Sigma yz + c\Sigma z^2$$

$$54 = 6a + 48b + 102c \quad \text{—— (1)}$$

$$339 = 48a + 494b + 1034c \quad \text{—— (2)}$$

$$720 = 102a + 1034b + 2188c \quad \text{—— (3)}$$

$$a = 16.47 \; ; \quad b = 0.3899 \; ; \quad c = -0.62$$

$$\boxed{x = 16.47 + 0.3899y - 0.62z}$$ **Ans.**

24. Calculate the rank correlation coefficient between X and Y from the following data-(CO1)

X	15	20	27	13	45	60	20	75
Y	50	30	55	30	25	10	30	70

Solution:

A	B	R(A)	R(B)	di	di^2
15	50	7	3	4	16
20	30	5.5	5	0.5	0.25
27	55	4	2	2	4
13	30	8	5	3	9
45	25	3	7	4	16
60	10	2	8	−6	36
20	30	5.5	5	0.5	0.25
75	70	1	1	0	0
					81.5

$$F = \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} = 2.5$$

$$\rho = 1 - \left[\frac{6(81.5+2.5)}{8(8^2-1)}\right]$$

$$= 1 - 1 = \underline{\underline{0}} \text{ Ans.}$$

25. The following results were obtained from record of age (x) and blood pressure (y) of a group of 10 men: (CO1)

	x	y
Mean	53	142
Variance	130	165

and $\sum(x-\bar{x})(y-\bar{y}) = 1220$

Find the appropriate regression equation and use it to estimate the blood pressure of a man whose age is 45.

Solution : $r = \dfrac{\sum(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}$.

'y on x': $(y-\bar{y}) = b_{yx}(x-\bar{x}) \Rightarrow b_{yx} = \dfrac{r\sigma_y}{\sigma_x}$

$\bar{x} = 53;\ \bar{y} = 142;\ \sigma_x^2 = 130;\ \sigma_y^2 = 165$

$r = \dfrac{1220}{10(\sqrt{130})(\sqrt{165})} = \dfrac{1220}{10(11.40)(12.84)}$

$= \dfrac{1220}{1463.76} = 0.833.$

$b_{yx} = \dfrac{r\sigma_y}{\sigma_x} = (0.833)\dfrac{\sqrt{165}}{\sqrt{130}} = 0.938$

$y - 142 = 0.938(x-53)$

$\boxed{x = 45}$

$y - 142 = 0.938(45-53)$

$y = -7.504 + 142$

$\boxed{y = 134.496}$ Ans.