Tutorial: Regression & Correlation

1. The data shows the respective heights x and y of a sample of 12 fathers and their oldest sons. (a) Construct a scatter diagram. (b) Find the least-squares regression line of y on x. (c) Find the least-squares regression line of x on y.

Height x of Father (inches)	65	63	67	64	68	62	70	66	68	67	69	71
Height y of Son (inches)	68	66	68	65	69	66	68	65	71	67	68	70

2. Find a least squares straight line for the following data:

X:	1	2	3	4	5	6
Y:	6	4	3	5	4	2

and estimate (predict) Y at X = 4 and X at Y = 4.

3. Consider a situation involving the number of blue refills X and the number of red refills Y is given. Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

		x		
	$f(x,y)$	0	1	2
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

Find the covariance of X and Y .Also, find the correlation coefficient between X and Y.

4. The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

Also, find the correlation coefficient between X and Y.

5. Given a random variable X, with standard deviation σ_X, and a random variable Y = a + bX, show that if b < 0, the correlation coefficient $\rho_{XY} = -1$, and if b > 0, $\rho_{XY} = 1$.
6. Random variables X and Y follow a joint distribution f(x, y) = 2 for $0 < x \le y < 1$, =0 otherwise. Determine the correlation coefficient between X and Y.
7. The following data give the resting pulse rates (in beats per minute) and the years of schooling of 10 individuals.

Person	1	2	3	4	5	6	7	8	9	10
Years of School	12	16	13	18	19	12	18	19	12	14
Pulse Rate	73	67	74	63	73	84	60	62	76	71

Find the correlation coefficient for these data, and analyse the high pulse rate's association with the number of years in school.

8. Determine rank correlation for the following data which shows the marks obtained in two quizzes in mathematics:

Marks in 1st quiz (X)	6	5	8	8	7	6	10	4	9	7
Marks in 2nd quiz (Y)	8	7	7	10	5	8	10	6	8	6