(Lect. 40)
11.12 Correlation

Up to this point we have assumed that the independent regressor variable x is a physical or scientific variable but not a random variable. In fact, in this context, x is often called a mathematical variable, which, in the sampling process, is measured with negligible error. In many applications of regression techniques, it is more realistic to assume that both X and Y are random variables and the measurements $(x_i, y_i); i = 1, 2, ..., n$ are observations from a population having the joint density function $f(x, y)$. We shall consider the problem of measuring the relationship between the two variables X and Y. For example, if X and Y represent the length and circumference of a particular kind of bone in the adult body, we might conduct an anthropological study to determine whether large values of X are associated with large values of Y, and vice versa.
Correlation analysis attempts to measure the strength of such relationships between two variables by means of a single number called a correlation coefficient.

Correlation coefficient

It is defined by

$$r = \frac{S_{xy}}{S_x S_y}$$

where S_x and S_y are standard deviation of x and y respectively.
S_{xy} is the covariance of xy.

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n}$$

$$E(x) = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$E(y) = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}}$$

Question No 43
Compute and interpret the correlation coefficient for the following grades of 6 students selected at random:

Mathematics grade	70	92	80	74	65	83
English grade	74	84	63	87	78	90

(0.1)

Solution: Let Mathematics grade=x
English grade=y

x	y	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
70	74	5180	53.7289	28.4089
92	84	7728	215.2089	21.8089
80	63	5040	7.1289	266.6689
74	87	6438	11.0889	58.8289
65	78	5070	152.0289	1.7689
83	90	7470	32.1489	113.8489
$\sum_{i=1}^{6} x_i = 464$	$\sum_{i=1}^{6} y_i = 476$	$\sum_{i=1}^{6} x_i y_i = 36926$	$\sum_{i=1}^{6} (x-\bar{x})^2 = 471.3334$	491.3334

(0.2)

$$E(x) = \bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = \frac{464}{6} = 77.33$$

$$E(y) = \bar{y} = \frac{1}{6}\sum_{i=1}^{6} y_i = \frac{476}{6} = 79.33$$

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{36926}{6} = 6154.33$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{471.3334}{5}} = 9.7091$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{491.3334}{5}} = 9.91295$$

$$S_{xy} = E(xy) - E(x)E(y) = E(xy) - \bar{x}\bar{y} = 6154.33 - (77.33 \times 79.33) = 19.4111$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{19.4111}{9.7091 \times 9.91295} = 0.2016825$$

Question No 45

A study of the amount of rainfall and the quantity of air pollution removed produced the following data.

$Daily Rainfall$ $x\ (0.01cm)$	$Particulate Removed,$ $y\ (\mu g/m3)$
4.3	126
4.5	121
5.9	116
5.6	118
6.1	114
5.2	118
3.8	132
2.1	141
7.5	108

(0.3)

assume a bivariate normal distribution for x and y. (a) Calculate r.

Solution

x	y	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
4.3	126	541.8	0.49	19.8025
4.5	121	544.5	0.25	0.3025
5.9	116	684.4	0.81	30.8025
5.6	118	660.8	0.36	12.6025
6.1	114	695.4	1.21	57.0025
5.2	118	613.6	0.04	12.6025
3.8	132	501.6	1.44	109.2025
2.1	141	296.1	8.41	378.3025
7.5	108	810	6.25	183.6025
$\sum_{i=1}^{9} x_i = 45$	$\sum_{i=1}^{9} y_i = 1094$	$\sum_{i=1}^{6} x_i y_i = 5348.2$	19.26	804.2225

(0.4)

$$E(x) = \bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{45}{9} = 5$$

$$E(y) = \bar{y} = \frac{1}{9}\sum_{i=1}^{9} y_i = \frac{1094}{9} = 121.55$$

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{5348.2}{9} = 594.244$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{19.26}{8}} = 1.5516$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{804.2225}{8}} = 10.0263$$

$$S_{xy} = E(xy) - E(x)E(y) = E(xy) - \bar{x}\bar{y} = 594.244 - (5 \times 121.55) = -13.506$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{-13.506}{1.5516 \times 10.0263} = -0.86817$$

Question No.47

The following data were obtained in a study of the relationship between the weight and chest size of infants at birth.

$Weight(kg)$	$ChestSize(cm)$
2.75	29.5
2.15	26.3
4.41	32.2
5.52	36.5
3.21	27.2
4.32	27.7
2.31	28.3
4.30	30.3
3.71	28.7

(0.5)

(a) Calculate r.

Solution

x	y	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
2.75	29.5	81.125	0.7744	0.0169
2.15	26.3	56.545	2.1904	11.0889
4.41	32.2	142.002	0.6084	6.6049
5.52	36.5	201.48	3.5721	47.1969
3.21	27.2	87.312	0.1764	5.9049
4.32	27.7	119.667	0.4761	3.7249
2.31	28.3	65.373	1.7424	1.7689
4.30	30.3	130.29	.4489	.4489
3.71	28.7	106.477	0.0064	0.8649
$\sum_{i=1}^{9} x_i = 32.68$	$\sum_{i=1}^{9} y_i = 266.7$	$\sum_{i=1}^{6} x_i y_i = 990.271$	9.9955	77.6201

(0.6)

$$E(x) = \bar{x} = \frac{1}{9}\sum_{i=1} x_i = \frac{32.68}{9} = 3.63$$

$$E(y) = \bar{y} = \frac{1}{9}\sum_{i=1}^{9} y_i = \frac{266.7}{9} = 29.63$$

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{990.271}{9} = 110.0301$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{9.9955}{8}} = 1.1177$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{77.6201}{8}} = 3.11488$$

$$S_{xy} = E(xy) - E(x)E(y) = E(xy) - \bar{x}\bar{y} = 110.0301 - (3.63 \times 29.63) = 2.4732$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{2.4732}{1.1177 \times 3.11488} = 0.710338$$