

Correlation Coefficient for grouped Data (in Arabic)

$$r = \frac{\sum xyf - N\bar{x}\bar{y}}{N * \sigma_x \sigma_y}$$

$$N = \sum f_i$$

where f_i is frequency

where:

x = mid points of class x

y = " " " " " " y

$$\bar{x} = \frac{\sum xf}{\sum f}, \quad \bar{y} = \frac{\sum yf}{\sum f}$$

σ_x & σ_y are standard deviation for x & y .

Example: Calculate the Correlation Coefficient by using Pearson's rule.

$x \backslash y$	15-25	25-35	35-45	45-55	$\sum fy$ (مجموع y في كل فئة x)
10-30	10	—	10	—	20
30-50	—	20	—	10	30
50-70	—	10	—	10	20
70-90	10	—	20	—	30
$\sum fx$ (مجموع x في كل فئة y)	20	30	30	20	100

So,

write a new table depend on the basic table after finding mid points of classes for x & y

الجدول الجديد يعتمد على الجدول الأساسي بعد إيجاد النقاط الوسطى للفئات x و y (4000) من فئة x * فئة y * في الجدول الجديد

مركز تكررات
 الفئات

x-classes	20	30	40	50	$\sum f_y$
20	4000	—	8000	—	12000
40	—	24000	—	20000	44000
60	—	18000	—	30000	48000
80	16000	—	64000	—	80000
$\sum f_x$	20000	42000	72000	50000	<u>184000</u>

class X	$F = \sum f_x$	X_i	$X_i f_i$	X_i^2	$X_i^2 f_i$
15-25	20	20	400	400	8000
25-35	30	30	900	900	27000
35-45	30	40	1200	1600	48000
45-55	20	50	1000	2500	50000
	$\sum = 100$		$\sum = 3500$		$\sum = 133000$

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{3500}{100} = 35$$

$$\sigma_x^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i} \right)^2 = \frac{133000}{100} - (35)^2 = 1330 - 1225 = 105$$

$$S_x = \sqrt{105} = 10.24$$

classes y	f_i	y_i	$y_i f_i$	y_i^2	$y_i^2 f_i$
10-30	20	20	400	400	8000
30-50	30	40	1200	1600	48000
50-70	20	60	1200	3600	72000
70-90	30	80	2400	6400	192000
	$\sum = 100$		$\sum = 5200$		$\sum = 320000$

$$\bar{y} = \frac{\sum y_i f_i}{\sum f_i} = \frac{5200}{100} = 52$$

$$S_y = \sqrt{\frac{\sum y_i^2 f_i}{\sum f_i} - \left(\frac{\sum y_i f_i}{\sum f_i} \right)^2} = \sqrt{3200 - 2704} = \sqrt{496} = 22.27$$

$$\therefore r = \frac{\sum xy - N\bar{x}\bar{y}}{N * S_x * S_y}$$

$$= \frac{184000 - 100 * (35)(52)}{100(10.24)(22.27)}$$

$$= 0.09$$

$$r = 0.09$$

Example: The following are the marks obtained by (132) students in two tests.

Test 1 \ Test 2	30-40	40-50	50-60	60-70	70-80	Total (1)
20-30	2	5	3			10
30-40	1	8	12	6		27
40-50		5	22	14	1	42
50-60		2	16	9	2	29
60-70		1	8	6	1	16
70-80			2	4	2	8
Total (2)	3	21	63	39	6	132

Calculate the Correlation Coefficient.

Rank Correlation [Spearman method]

This method is based on ranks. It's useful to study the qualitative measure of attributes like honesty, colour, beauty, morality.

$$r = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \text{ where } d_i = \text{rank } X_i - \text{rank } Y_i$$

Example: In a marketing survey the price of tea and coffee in a town based on quality was found as shown below. Could you find any relation between tea & coffee price.

Price of tea	88	90	95	70	60	75	50
Price of Coffee	120	134	150	115	110	140	100

Sol:

Price of T	Price of C	Rank T	Rank C	D	D ²
88	120	3	4	1	1
90	134	2	3	1	1
95	150	1	1	0	0
70	115	5	5	0	0
60	110	6	6	0	0
75	140	4	2	2	4
50	100	7	7	0	0
					$\Sigma = 6$

$$r = 1 - \frac{6 \sum d^2}{n^3 - n} = 1 - \frac{6 \times 6}{7^3 - 7}$$

$$= 1 - \frac{36}{336} = 1 - 0.1071 = 0.8929$$

In the past example notice that the relation between price of tea and coffee is positive at 0.89
Based on quality the association between price of tea and price of coffee is highly positive.

Example 2: The following data represent marks (تقديرات) for seven students in Math & Statistics.
Could you find any relation between them.

Math good Poor Excellent good v-good poor good
Stati. v-good Poor v-good good good good Excellent

Sol: ترتيب التقديرات بترتيب تنازلي
 $X_i =$ ① Poor ② Poor ③ good ④ good ⑤ good ⑥ v-good ⑦ Excellent
 $Y_i =$ ① poor ② good ③ good ④ good ⑤ v-good ⑥ v-good ⑦ Excellent

Math. (X)	Stati. (Y)	rank(X)	rank(Y)	d	d ²
good	v-good	4	5.5		
Poor	Poor	1.5	1		
Exce.	v-good	7	5.5		
good	good	4	3		
v-good	good	6	3		
Poor	good	1.5	3		
good	EXCe	4	7		

(بالترتيب التنازلي نأخذ الرقم 1)
 $u =$ 3, 4, 5
 good بالترتيب 4
 وبالترتيب رقم 4