27/9/21

MAT 6001

CORRELATION (contd)

Problem 2

Example:

The following data represents the number of hours 12 different students watched television during the weekend and the scores of each student who took a test the following Monday.

Calculate the correlation coefficient *r*.

Hours, x	0	1	2	3	3	5	5	5	6	7	7	10
Test score, y	96	85	82	74	95	68	76	84	58	65	75	50

Problem 2 continued:

$n = 12$

Hours, x	0	1	2	3	3	5	5	5	6	7	7	10
Test score, y	96	85	82	74	95	68	76	84	58	65	75	50
xy	0	85	164	222	285	340	380	420	348	455	525	500
x^2	0	1	4	9	9	25	25	25	36	49	49	100
y^2	9216	7225	6724	5476	9025	4624	5776	7056	3364	4225	5625	2500

$\sum x = 54 \qquad \sum y = 908 \qquad \sum xy = 3724 \qquad \sum x^2 = 332 \qquad \sum y^2 = 70836$

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{12(3724) - (54)(908)}{\sqrt{12(332) - 54^2}\sqrt{12(70836) - (908)^2}} \approx -0.831$$

There is a strong negative linear correlation.

As the number of hours spent watching TV increases, the test scores tend to decrease.

Problem 3

- The following are the marks scored by 7 students in two tests in a subject. Calculate coefficient of correlation from the following data and interpret.

Marks in test-1	12	9	8	10	11	13	7
Marks in test-2	14	8	6	9	11	12	3

- Let x denote height of father and y denote height of son. The data is on the ratio scale.
- We use Karl Pearson's method.

x_i	y_i	x_i^2	y_i^2	xy_i
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
544	552	37028	38132	37560

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

$$r = \frac{8 \times 37560 - 544 \times 552}{\sqrt{8 \times 37028 - (544)^2}\sqrt{8 \times 38132 - (552)^2}} = 0.603$$

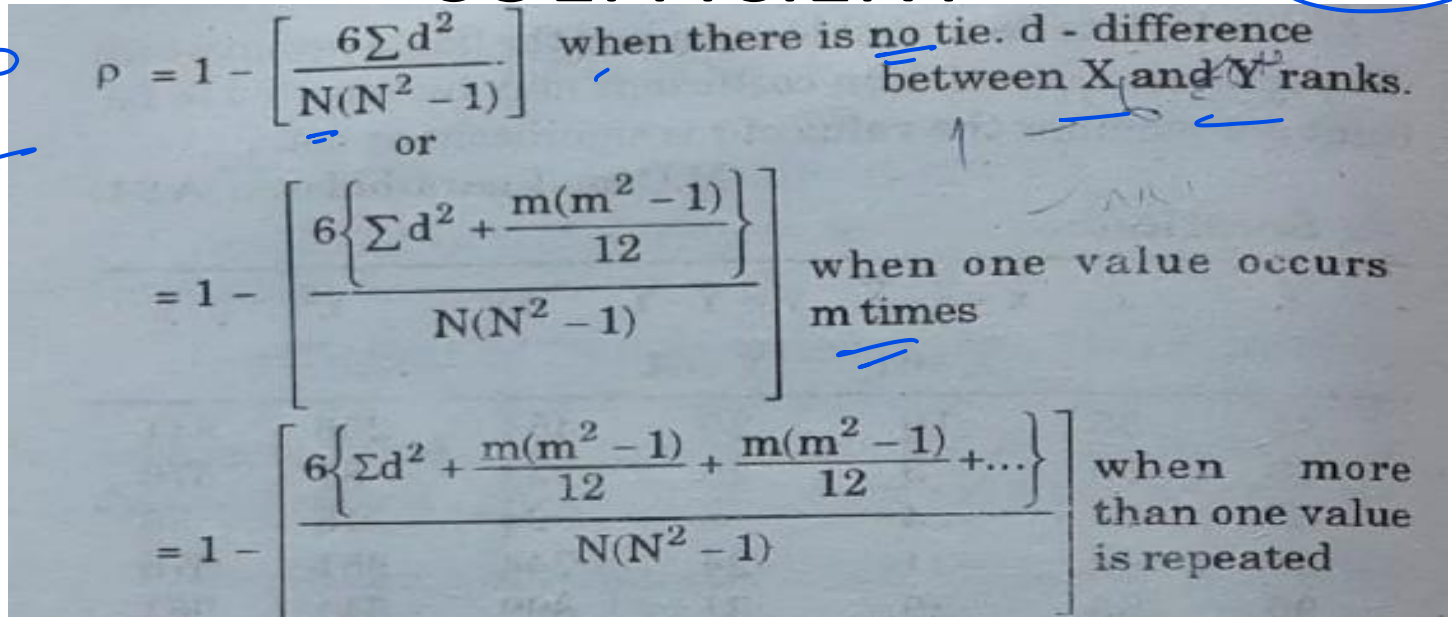
- Heights of father and son are positively correlated. It means that on the average , if fathers are tall then sons will probably tall and if fathers are short, probably sons may be short.

Extra problems

- A computer while calculating r_{xy} form 25 pairs of observaions obtained the following constants
  n=25
  $\sum x = 125, \sum x^2 = 650, \sum y = 100, \sum y^2 = 460, \sum xy = 508$
- A recheck showed that 2 pairs of values (6,14) and (8.6) were wrong while the correct values were (8,12) and (6,8). Obtain the correct correlation coefficient.
- Solution:
- Corrected values
- $\sum x = 125, \sum x^2 = 650, \sum y = 100, \sum y^2 = 436, \sum xy = 520$

- r_{xy}=0.667

SPEARMANS RANK CORRELATION COEFFICIENT



$$\rho = 1 - \left[\frac{6\sum d^2}{N(N^2-1)}\right]$$ when there is no tie. d - difference between X and Y ranks.

or

$$= 1 - \left[\frac{6\left\{\sum d^2 + \frac{m(m^2-1)}{12}\right\}}{N(N^2-1)}\right]$$ when one value occurs m times

$$= 1 - \left[\frac{6\left\{\Sigma d^2 + \frac{m(m^2-1)}{12} + \frac{m(m^2-1)}{12} + \ldots\right\}}{N(N^2-1)}\right]$$ when more than one value is repeated

- It is calculated when ranks are given or when rank correlation coefficient is required. Rank correlation lies between -1 and 1

problems

Example 1: Rankings of 10 trainees at the beginning (X) and at the end (Y) of a certain course are given below:

Trainees :	A	B	C	D	E	F	G	H	I	J
X	1	6	3	9	5	2	7	10	8	4
Y	6	8	3	7	2	1	5	9	4	10

Calculate Spearman's rank correlation coefficient

Solution :

X	Y	d	d^2
1	6	−5	25
6	8	−2	4
3	3	0	0
9	7	2	4
5	2	3	9
2	1	1	1
7	5	2	4
10	9	1	1
8	4	4	16
4	10	−6	36
Total	---	$\sum d = 0$	$\sum d^2 = 100$

$$\rho = 1 - \left[\frac{6\sum d^2}{N(N^2-1)}\right]$$

$$= 1 - \left[\frac{6 \times 100}{10 \times 99}\right]$$

$$= 1 - 0.6061$$

$$= 0.3939$$

Example : Marks obtained by 8 students in Accountancy (X) and Statistics (Y) are given below. Compute rank correlation.

X	15	20	28	12	40	60	20	80
Y	40	30	50	30	20	10	30	60

Solution:

X	Y	Ranks X	Ranks Y	d	d^2
15	40	7	3	4	16
20	30	5.5	5	0.5	0.25
28	50	4	2	2	4
12	30	8	5	3	9
40	20	3	7	−4	16
60	10	2	8	−6	36
20	30	5.5	5	0.5	0.25
80	60	1	1	0	0
Total	—	—	—	$\sum d = 0$	$\sum d^2 = 81.5$

$$\rho = 1 - \left[\frac{6\left\{\sum d^2 + \frac{m(m^2-1)}{12} + \frac{m(m^2-1)}{12}\right\}}{N(N^2-1)}\right]$$

$$= 1 - \left[\frac{6\{81.5 + 0.5 + 2\}}{8(8^2-1)}\right]$$

$$= 1 - \left[\frac{6 \times 84}{8 \times 63}\right]$$

$$= 0$$

Note :

Item	Freq.	Probable Ranks	Rank Assigned	$\frac{m(m^2-1)}{12}$
X = 20	m = 2	5, 6	$\frac{5+6}{2} = 5.5$	$\frac{2(2^2-1)}{12} = 0.5$
Y = 30	m = 3	4, 5, 6	$\frac{4+5+6}{3} = 5.0$	$\frac{3(3^2-1)}{12} = 2.0$

30 → 4th, 5th, 6th

d = diff of x & y

Conclusion:-

conclusion from the problem.

" " scenario.

→ marks