



*Dr. Bhasa Science*


Sponsored by  
**The Science Foundation College**  
**Uganda East Africa**  
Senior one to senior six

**+256 778 633682 0753 143413**

**Based on, Best for Science**

[digitalteachers.co.ug](http://digitalteachers.co.ug)



Nuture your dreams 

## SENIOR FIVE TERM 2

### TOPIC 1/6: Scatter Diagrams and Correlations

**Competency:** The learner investigates the relationship between two quantitative variables through graphical representation and evaluation of correlation coefficients for decision making and prediction.

#### Contents

Scatter diagrams and Correlations.....	2
Interpretation of the rank correlation coefficients.....	2
Spearman rank correlation ( $\rho$ ).....	2
Kendall's Rank correlation coefficient .....	4
Significance of ranks correlation coefficients .....	6
Scatter graphs .....	7
Revision questions on correlation and scatter diagrams.....	10

## Scatter diagrams and Correlations

Correlation refers to **the degree of correspondence or relationship between two variables**. Correlated variables tend to change together. If one variable gets larger, the other one systematically becomes either larger or smaller.

The degree of correlation/association is determined by rank correlation coefficients

There are two types

1. Spearman rank correlation coefficient ( $\rho$ )
2. Kendall's rank correlation coefficient ( $\tau$ )

### Interpretation of the rank correlation coefficients

A rank correlation coefficient measures the degree of similarity between two rankings

The table below is used

Correlation coefficient	Interpretation
0-0.19	Very low correlation
0.2-0.39	Low correlation
0.4-0.59	Moderate correlation
0.6 – 0.79	High correlation
0.8 – 1.0	Very high correlation

Note: the positive or negative signs indicate positive or negative relationships respectively. Or they the relationships are directly or inversely related.

The closer to zero the lower the relationship

### Spearman rank correlation ( $\rho$ )

It is given by  $\rho = 1 - \left[ \frac{6 \sum d^2}{n(n^2-1)} \right]$

Where  $d$  = difference between ranks

$n$  = total number of pairs

#### Example 1

Two examiners marked the scripts of 8 candidates. The table shows the marks awarded by two examiners  $x$  and  $y$ .

$x$	72	60	56	76	68	52	80	64
$y$	56	44	60	74	66	38	68	52

Please find free notes, tests and marking guides of physics, chemistry, biology and mathematics from: [digitalteachers.co.ug](http://digitalteachers.co.ug)

Calculate the rank correlation coefficient and comment on your results

Solution

R <sub>x</sub>	R <sub>y</sub>	d	d <sup>2</sup>
3	5	2	4
6	7	1	1
7	4	3	9
2	1	1	1
4	3	1	1
8	8	0	0
1	2	-1	1
5	5	1	1
			$\sum d^2 = 18$

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

$$= 1 - \left[ \frac{6 \times 18}{8(8^2 - 1)} \right] = 0.786$$

There is a high positive correlation between x and y

### Example 2

The following shows the marks obtained by 10 students in mathematics and physics exams.

Mathematics	80	80	70	60	65	80	68	90	95	50
Physics	50	45	70	80	70	90	70	80	70	95

Calculate the ranks correlation coefficient and comment on your results

Solution

R <sub>M</sub>	R <sub>P</sub>	d	d <sup>2</sup>
4	9	-5	25
6	6.5	0.5	0.25
9	3.5	5.5	30.25
8	6.5	1.5	2.25
4	2	2	4
7	6.5	0.5	0.25
2	3.5	1.5	2.25
1	6.5	-5.5	30.25
10	1	9	81
			$\sum d^2 = 211.5$

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

$$= 1 - \left[ \frac{6 \times 211.5}{10(10^2 - 1)} \right] = -0.282$$

There is a low negative correlation between mathematics and physics

### Example 3

The following table gives the order in which six candidates were ranked in two tests x and y

x	E	C	B	F	D	A
y	F	A	D	E	C	C

Calculate the rank correlation coefficient and comment on your results

Solution

R <sub>x</sub>	R <sub>y</sub>	d	d <sup>2</sup>
5	6	1	1
3	1	2	4
2	4	2	4
6	5	1	1
4	2.5	1.5	2.25
1	2.5	-1.5	2.25
			$\sum d^2 = 14.5$

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

$$= 1 - \left[ \frac{6 \times 14.5}{6(6^2 - 1)} \right] = -0.586$$

There is a moderate negative correlation between x and y

### Kendall's Rank correlation coefficient

It is a coefficient that represents the degree of concordance/agreement between two columns of ranked data. The greater the 'inversions' the smaller the coefficient will

$$\text{Kendall's Rank correlation coefficient, } \tau (\text{tau}) = \frac{C - D}{C + D}$$

Where C = number of concordant pairs or pairs in agreement

D = number of discordant pairs or pairs in disagreement

The Tau correlation coefficient returns a value of 0 to 1, where: 0 is no relationship, 1 is a perfect relationship

- Concordant pairs are the number of observed ranks below a particular rank which are larger than that particular rank.
- Discordant pairs are the number of observed rank below a particular rank which are smaller than that particular rank

### Example 4

Two examiners marked the scripts of 8 candidates. The table shows the marks awarded by two examiners x and y.

x	72	60	56	76	68	52	80	64
y	56	44	60	74	66	38	68	52

Calculate the rank correlation coefficient and comment on your results

Solution

Ranking values

The subscripts are the ranks

x	72 <sub>3</sub>	60 <sub>6</sub>	56 <sub>7</sub>	76 <sub>2</sub>	68 <sub>4</sub>	52 <sub>8</sub>	80 <sub>1</sub>	64 <sub>5</sub>
y	56 <sub>5</sub>	44 <sub>7</sub>	60 <sub>4</sub>	74 <sub>1</sub>	66 <sub>3</sub>	38 <sub>8</sub>	68 <sub>2</sub>	52 <sub>6</sub>

Please find free notes, tests and marking guides of physics, chemistry, biology and mathematics from: [digitalteachers.co.uk](http://digitalteachers.co.uk)

- The ranks of x and y are filled in the table as below those of x in ascending order and those of y correspondingly as shown in the table below.
- The values of C are bigger values in the column Ry bigger than and below a particular value in that column. While the values of D are smaller values in the column Ry bigger than and below a particular value in that column
- For instance, the first value of C in the table below is the number of values bigger than and below 2 in column Ry i.e. (5, 3, 6, 7, 4, 8) =6; the first value D = the number of values smaller than and below 2 in column Ry; i.e. (1) = 1

Rx	Ry	C	D
1	2	6	1
2	1	6	0
3	5	3	2
4	3	3	0
5	6	2	1
6	7	1	1
7	4	1	0
8	8		
		$\sum C = 22$	$\sum D = 5$

$$\tau (\text{tau}) = \frac{C-D}{C+D}$$

$$\tau (\text{tau}) = \frac{22-5}{22+5} = 0.63$$

moderate positive correlation

### Example 5

The height (cm) and ages (years) of random sample of ten farmers are given in the table below

Height (cm)	156	151	152	160	146	157	149	142	158	140
Ages (years)	47	38	44	55	46	49	54	52	45	30

- (a)(i) Calculate the Kendall rank correlation coefficient  
(ii) comment on your result (06marks)

Let the farmers be A, B, C, D, E, F, G, H, I, J (the subscripts are the rank)

Farmers	A	B	C	D	E	F	G	H	I	J
Height	156 <sub>4</sub>	151 <sub>6</sub>	152 <sub>5</sub>	160 <sub>1</sub>	146 <sub>8</sub>	157 <sub>3</sub>	149 <sub>7</sub>	142 <sub>9</sub>	158 <sub>2</sub>	140 <sub>10</sub>
Age	47 <sub>5</sub>	38 <sub>9</sub>	44 <sub>8</sub>	55 <sub>1</sub>	46 <sub>6</sub>	49 <sub>4</sub>	54 <sub>2</sub>	52 <sub>3</sub>	45 <sub>7</sub>	30 <sub>10</sub>

By re-arranging the findings we have

Farmers	D	I	F	A	C	B	G	E	H	J
Height	1	2	3	4	5	6	7	8	9	10
Age	1	7	4	5	8	9	2	6	3	10
Agreements (C)	9	3	5	4	2	1	3	1	1	=29
Disagreements (D)	0	4	2	2	3	3	0	1	0	=15

$$\tau (\text{tau}) = \frac{C-D}{C+D}$$

$$\tau (\text{tau}) = \frac{29-15}{29+15} = 0.32$$

low positive correlation

### Significance of ranks correlation coefficients

The calculated ranks correlation coefficients to be statistically significant; the calculated value should be greater than that from the table critical values associated with the various sample sizes and significance levels ( $\alpha$ ).

That is

- ⇒ If the  $|\rho_c| > |\rho_T|$  or  $|\tau_c| > |\tau_T|$ , a significant relationship exists
- ⇒ If the  $|\rho_c| < |\rho_T|$  or  $|\tau_c| < |\tau_T|$ , no significant relationship exists

Where  $\rho_c$  = calculated Spearman's correlation coefficient

$\rho_T$  = Table Spearman's correlation coefficient at either 1% ( $\alpha=0.01$ ) or 5% ( $\alpha=0.05$ )

$\tau_c$  = calculated Kendall's correlation coefficient

$\tau_T$  = Table Kendall's correlation coefficient at either 1% ( $\alpha=0.01$ ) or 5% ( $\alpha=0.05$ )

### Example 6

The following shows the marks obtained by 8 students in mathematics and physics exams

Mathematics	65	65	70	75	75	80	85	85
Physics	50	55	58	55	65	58	61	65

Calculate the ranks correlation coefficient and comment of the significance of your results at 5% level (Spearman's  $\rho = 0.71$ ), Kendall's  $\tau = 0.64$

(a) Using Spearman's correlation coefficient

$R_M$	$R_p$	$d$	$d^2$
7.5	8	0.5	0.25
7.5	6.5	1	1
6	4.5	1.5	2.25
4.5	6.5	-2	4
4.5	1.5	3	9
3	4.5	-1.5	2.25
1.5	3	-1.5	2.25
1.5	1.5	0	0
			$\sum d^2 = 21$

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

$$= 1 - \left[ \frac{6 \times 21}{8(8^2 - 1)} \right] = 0.75$$

Since  $\rho_c(0.75) > \rho_T(0.71)$ , a significant relationship exist

Using Kendall's correlation coefficient

$R_M$	$R_p$	C	D
1.5	1.5	6	0
1.5	3	5	1
3	4.5	3	1
4.5	1.5	4	0
4.5	6.5	1	1
6	4.5	2	0
7.5	6.5	1	0
7.5	8	=22	3

$$\tau(\text{tau}) = \frac{C - D}{C + D}$$

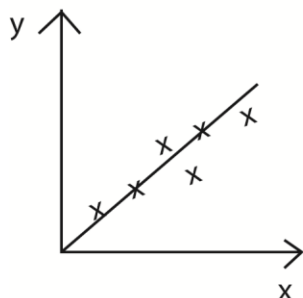
$$\tau(\text{tau}) = \frac{22 - 3}{22 + 3} = 0.76$$

Since  $\tau_c(0.75) > \tau_T(0.64)$ , a significant relationship exist

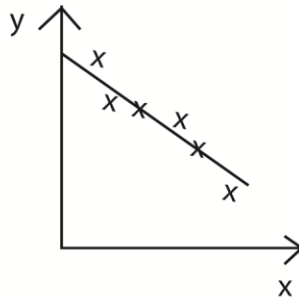
## Scatter graphs

They are graphs showing the relationship between two variables

Positive relationship



Negative relationship



### Example 7

The heights and masses of ten students are given in the table below

Height (cm)	156	152	152	146	160	157	149	142	158	68
Mass (kg)	62	58	63	58	70	60	55	57	68	56

(a)(i) Plot the data on a scatter diagram

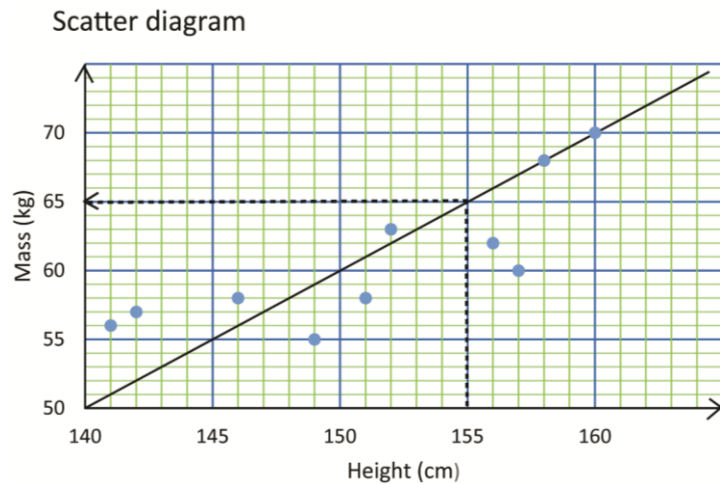
(ii) Draw the line of the best fit. Hence estimate the mass corresponding to height of 155cm

(b) (i) Calculate the rank correlation coefficient for the data.

(ii) Comment on the significance of the heights on the masses of the students. [Spearman's  $\rho = 0.79$  and Kendall's  $\tau = 0.64$  at 1% level of significance based on 10 observations]

Solution

(a)(i)



The weight corresponding to height 155cm is 65kg

**Note:** this value may vary from 63 to 67kg depending on how one has drawn the line of the best fit.

**(b)(i) Using Spearman's rank correlation coefficient**

Heights (x)	Mass (y)	R <sub>x</sub>	R <sub>y</sub>	d	d <sup>2</sup>
156	62	4	4	0	0
151	58	6	6.5	0.5	0.25
152	63	5	3	2	4
146	58	8	6.5	1.5	2.25
160	70	1	1	0	0
157	60	3	5	-2	4
149	55	7	10	-3	9
142	57	9	8	1	1
158	68	2	2	0	0
141	56	10	9	1	1
					$\sum d^2 = 21.5$

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

$$= 1 - \left[ \frac{6 \times 21.5}{10(10^2 - 1)} \right] = 0.8697$$

Since  $\rho_c(0.87) > \rho_T(0.79)$ , a significant relationship exist between height and weight of students at 1% level

Using Kendall's method

By naming the pairs we have

A(156, 62), B(151, 58), C(152, 63), D(146, 58), E(160, 70), F(157, 60), G(149, 55), H(142, 57), I(158, 68), J(141, 56)

	E	I	F	A	C	B	G	D	H	J
x	1	2	3	4	5	6	7	8	9	10
y	1	2	5	4	3	6.5	10	6.5	8	9
C	9	8	5	5	5	3	0	2	1	=38
D	0	0	2	1	0	0	3	0	0	=6

$$\tau(\text{tau}) = \frac{C-D}{C+D}$$

$$\tau(\text{tau}) = \frac{38-6}{38+6} = 0.73$$

Since  $\tau_c(0.73) > \tau_T(0.64)$ , a significant relationship exist between the heights and masses of student.

## Revision questions on correlation and scatter diagrams

1. Eight candidate seeing admission to a University sat for written and oral test. The scores were shown below

Written (x)	55	54	35	62	87	53	71	50
Oral (y)	57	60	47	65	83	56	74	63

- (a) Plot the result on a scatter diagram. Comment on the relationship between the written test and oral test
- (b) Draw the line of the best fit on your graph and use it to estimate y when x = 70.
- (c) Calculate the rank correlation coefficient. Comment on your results
2. The pairs of observation have been made on two random variables x and y. the ten (x, y) are (0, 20), (-7, 12), (-10, 15), (-12, 22), (-17, 5), (-30, -5), (-32, 13), (10, 30), (15, 40), and (-12, 8).
- (a) Draw the results on a scatter diagram
- (b) Draw the line of the best fit
- (c) Estimate the expected value of y corresponding to x = -7
- (d) Calculate the rank correlation coefficient and comment on the significance of the results at 1% significance level. ( $\rho = 0.894$ ,  $\tau = 0.778$ )
3. Three examiners X, Y and Z each marked scripts of ten candidates who sat for mathematics examination. The table below shows the examiner's ranking of candidates.

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

Calculate the coefficient of rank correlation of the rankings

- (i) X and Y
- (ii) Y and Z
- (iii) Comment on the significance of each at 5% significant level
4. Three weighing scales from three different shops W, X and Y in a market were used to weigh 10 bags of beans (A, B, C.....) and the results in (kg) were given in the table below

	A	B	C	D	E	F	G	H	I	J
W	65	68	70	63	64	62	73	75	72	78
X	63	68	68	60	65	60	72	73	70	66
Y	63	74	78	75	64	73	79	70	67	79

Determine the rank correlation coefficient for the performance of the scales

- (i) W and X
- (ii) X and Y
- (iii) Which of the three scales W, X and Y were in good working conditions
5. (a) In many government institution, officers complain about typing errors. A test was designed to investigate the relationship between typing speed and errors made. Twelve typist A, B, C ...L were picked at random to type a text. The table below shows the rankings of the typist according to speed and errors made. (N.B lowest ranking in error implies the least errors

Typist	A	B	C	D	E	F	G	H	I	J	K	L
Speed	3	4	2	1	8	11	10	6	7	12	5	9
Errors	2	6	5	1	10	9	8	3	4	12	7	11

- (i) Calculate the coefficient of rank correlation  
(ii) Comment on the significance at 1% significance level
- (b) The cost of travelling at a certain distance away from the city centre is found to depend on the route and distance a given place is away from the centre. The table below gives average rates of travel charged for distances to be travelled away from the city centre

Distance (s km)	9	12	14	21	24	30	33	45	46	50
Rate charged (r shs)	750	1000	1150	1200	1350	1250	1400	1750	1600	2000

- (i) Plot the above data on a scatter diagram and draw a line of best fit through the points of the scatter diagram  
(ii) Estimate the expected value r corresponding to s = 40km

6. In a certain commercial institution, a speed and error typing examination was administered to 12 randomly selected candidates A, B, C, ..., L of the institution. The table below shows their speed (y) in seconds and the number of errors in their typed scripts (x)

	A	B	C	D	E	F	G	H	I	J	K	L
No. of errors (x)	12	24	20	10	32	30	28	15	18	40	27	35
Speed (y) in seconds	130	136	124	120	153	160	155	142	145	172	140	157

- (i) Calculate the coefficient of rank correlation of the ranking  
(ii) Comment on your results  
(iii) Plot the above data on a scatter diagram and draw a line of the best fit through the points of the scatter diagram

7. The following table gives the marks obtained in calculus, physics and statistics by seven students

Calculus	72	50	60	55	35	48	82
Physics	61	55	70	50	30	50	73
Statistics	50	40	62	70	40	40	60

Determine the rank correlation coefficient for the performance of students in

- (i) Calculus and physics  
(ii) Calculus and statistics
8. Given the table below

x	80	75	86	60	75	92	86	50	64	75
y	62	58	60	45	68	68	81	48	50	70

Determine the rank correlation coefficient between the variable x and y, comment on your results

9. The table below shows the percentage of sand y in the soils at different depth x (in cm)

Soil depth (x)(cm)	35	65	55	25	45	75	20	90	51	60
% of sand, y	86	70	84	92	79	68	96	58	86	77

- (a) Plot the results on a scatter diagram. Comment on the relationship between the depth of the soil and the percentage of sand in the soil
- (b) Draw the line of the best fit on you graph and use it to estimate
- The percentage of sand in the soil at a depth of 31cm
  - Depth of the soil with 54% sand
  - Calculate the rank correlation coefficient

10. Eight applicants for a certain job obtained the following marks in aptitude and written test

Applicants	A	B	C	D	E	F	G	H
Aptitude test	33	45	16	42	45	35	40	48
Written test	57	60	40	75	68	48	54	68

- Calculate the coefficient of rank correlation of applicant's performance in the two tests
- Comment on your results

11. The table below shows the marks scored by students in mathematics and fine art tests

students	A	B	C	D	E	F	G	H	I	J
Mathematics,	40	48	79	26	55	35	37	70	60	40
Fine art	59	62	68	47	46	39	63	29	55	67

Calculate the coefficient of rank correlation for the students' performance in the two subjects and comment on your results.

12. Below are marks scored by 8 students A, B, C ... H in Mathematics, Economics and geography in the end of term examination.

	A	B	C	D	E	F	G	H
Math	52	75	41	60	81	31	65	52
Economics	50	60	35	65	66	45	60	48
Geography	59	62	68	54	63	40	55	72

Determine the rank correlation coefficient for the performance of students in

- Math and economics
- Geography and math
- Comment on the significance of the math in performance of economics and geography. ( $\rho = 0.86$ ,  $\tau = 0.79$  based on 8 observations at a 1% level of significance)

13. The heights and ages of ten students are given in the table below

Height, cm	156	151	152	146	160	157	149	142	158	140
Mass, kg	62	58	63	58	70	60	55	57	68	56

- Plot the data on a scatter diagram
- Draw the line of best fit on you graph and use it to estimate the mass corresponding to a height of 155cm
- Calculate the rank correlation coefficient for the data. Comment on the significance of the height on masses of students ( $\rho = 0.79$ ,  $\tau = 0.64$  based on 10 observations at 1% level of significance).

14. The heights and ages of ten farmers are given in the table below

Height, cm	156	151	152	160	146	157	149	142	158	140
------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Age, years	47	38	44	55	46	49	45	30	45	20
------------	----	----	----	----	----	----	----	----	----	----

- (a) Plot the data on a scatter diagram  
 (b) Draw the line of best fit on your diagram and use it to estimate  
 (i) Age when height = 147  
 (ii) Height when the age is 43  
 (c) Calculate the rank correlation coefficient for the data. Comment on your results ( $\rho=0.752$ ,  $\tau=0.6$ )
15. The table gives the points awarded to eight schools by three judges,  $J_1$ ,  $J_2$  and  $J_3$  during a music competition.  $J_1$  was the chief judge.

$J_1$	72	50	50	55	35	38	82	72
$J_2$	60	55	70	50	50	50	73	70
$J_3$	50	40	62	70	40	48	67	67

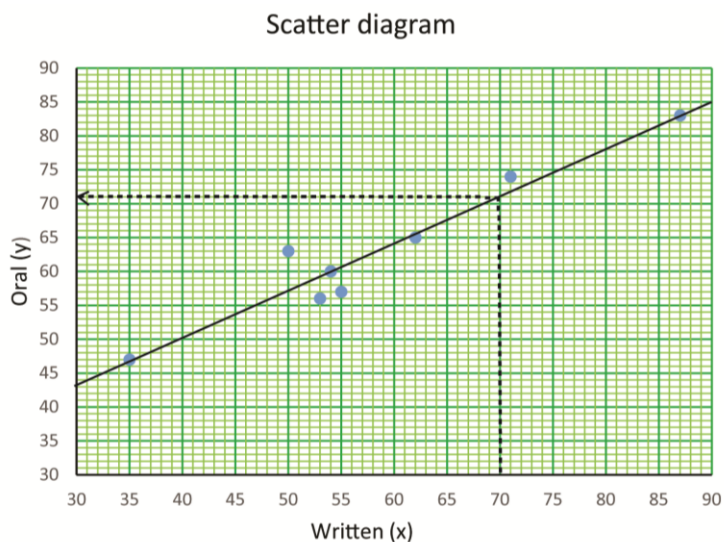
- (a) Determine the rank correlation coefficients between the judges of  
 (i)  $J_1$  and  $J_2$   
 (ii)  $J_1$  and  $J_3$   
 (b) Who of the two judges had a better correlation with the chief judge? Give a reason.

### Answers to the revision exercise

1. Eight candidate seeing admission to a University sat for written and oral test. The scores were shown below

Written (x)	55	54	35	62	87	53	71	50
Oral (y)	57	60	47	65	83	56	74	63

- (a) Plot the result on a scatter diagram. Comment on the relationship between the written test and oral test



- (b) Draw the line of the best fit on your graph and use it to estimate y when x = 70. The value of y when x = 70 is 71
- (c) Calculate the rank correlation coefficient. Comment on your results

Using Spearman's rank correlation method

x	y	R <sub>x</sub>	R <sub>y</sub>	d	d <sup>2</sup>
55	57	4	6	-2	4
54	60	5	5	0	0
35	47	8	8	0	0
62	65	3	3	0	0
87	83	1	1	0	0
53	56	6	7	-1	1
71	74	2	2	0	0
50	63	7	4	3	9
					$\sum d^2 = 14$

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

$$= 1 - \left[ \frac{6 \times 14}{8(8^2 - 1)} \right] = 0.833$$

There is very high correlation between x and y

Using Kendall's Rank correlation coefficient method

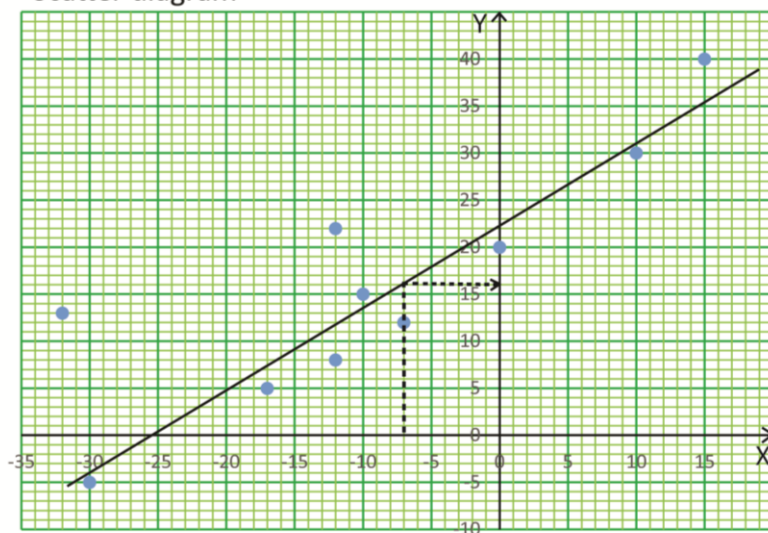
R <sub>x</sub>	R <sub>y</sub>	C	D
1	1	7	0
2	2	6	0
3	3	5	0
4	6	3	1
5	5	2	1
6	7	1	1
7	4	1	0
8	8	=25	=3

$$\tau = \frac{C-D}{C+D} = \frac{25-3}{25+3} = 0.786$$

there is high correlation between x and y

2. The pairs of observation have been made on two random variables x and y. the ten (x, y) are (0, 20), (-7, 12), (-10, 15), (-12, 22), (-17, 5), (-30, -5), (-32, 13), (10, 30), (15, 40), and (-12, 8).
- (a) Draw the results on a scatter diagram
- (b) Draw the line of the best fit

Scatter diagram



- (c) Estimate the expected value of y corresponding to x = -7  
The value of y corresponding to x = -7 is 15.7
- (d) Calculate the rank correlation coefficient and comment on the significance of the results at 1% significance level. ( $\rho = 0.894$ ,  $\tau = 0.778$ )  
Using Spearman's rank correlation coefficient

X	Y	R <sub>x</sub>	R <sub>y</sub>	d	d <sup>2</sup>
0	20	3	4	-1	1
-7	12	4	7	-3	9
-10	15	5	5	0	0
-12	22	6.5	3	3.5	12.25
-17	5	8	9	-1	1
-30	-5	9	10	-1	1
-32	13	10	6	4	16
10	30	2	2	0	0
15	40	1	1	0	0
-12	8	6.5	8	-1.5	2.25
					$\sum d^2 = 42.5$

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

$$= 1 - \left[ \frac{6 \times 42.5}{10(10^2 - 1)} \right] = 0.742$$

Since  $\rho_c (0.742) < \rho_\tau (0.894)$

There is no significant relationship between X and Y at 1% significance level

Using Kendall's rank correlation coefficient

Rx	Ry	C	D
1	1	9	0
2	2	8	0
3	4	6	1
4	7	3	3
5	5	4	1
6.5	3	4	0
6.5	8	2	1
8	9	1	1
9	10	0	1
10	6	=37	=8

$$\tau = \frac{C-D}{C+D} = \frac{37-8}{37+8} = 0.644$$

Since  $\tau_c (0.644) < \tau_T (0.778)$

There is no significant relationship between X and Y at 1% significance level.

3. Three examiners X, Y and Z each marked scripts of ten candidates who sat for mathematics examination. The table below shows the examiner's ranking of candidates.

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

Calculate the coefficient of rank correlation of the rankings

- (i) X and Y

Spearman's rank correlation

Rx	Ry	d	d <sup>2</sup>
8	5	3	9
5	3	2	4
9	6	3	9
2	1	1	1
10	4	6	36
1	7	-6	36
7	2	5	25
6	10	-4	16
3	8	-5	25
4	9	-5	25
			$\sum d^2 = 186$

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

$$= 1 - \left[ \frac{6 \times 186}{10(10^2-1)} \right] = -0.127$$

There is low negative relationship between X and Y

Kendall's rank correlation coefficient

Rx	Ry	C	D
1	7	3	6
2	1	8	0
3	8	2	5
4	9	1	5
5	3	4	1
6	10	0	4
7	2	3	0
8	5	1	1
9	6	0	1
10	4	=22	=23

$$\tau = \frac{C-D}{C+D} = \frac{22-23}{22+23} = -0.022$$

There is very low negative relationship between X and Y at 1% significance level.

(ii) Y and Z

Spearman's rank correlation coefficient

Ry	Rz	d	d <sup>2</sup>
5	6	-1	1
3	3	0	0
6	7	-1	1
1	2	-1	1
4	5	-1	1
7	4	3	9
2	1	1	1
10	10	0	0
8	9	-1	1
9	8	1	1
			$\sum d^2 = 16$

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

$$= 1 - \left[ \frac{6 \times 16}{10(10^2-1)} \right] = 0.903$$

There is high positive relationship between Y and Z

Kendall's rank correlation coefficient

Rz	Ry	C	D
1	2	9	1
2	1	8	0
3	3	7	0
4	7	3	3
5	4	5	0
6	5	4	0
7	6	3	0
8	9	1	1
9	8	1	0
10	10	41	5

$$\tau = \frac{C-D}{C+D} = \frac{41-5}{41+5} = 0.7826$$

There is very high relationship between X and Y

(iii) Comment on the significance of each at 5% significant level  
(From the tables of critical values at 5% level of significance based on 10 observations,  $\rho = 0.648$ ,  $\tau = 0.467$ )

(i) For X and Y since  $|\rho_C(-0.127)| < |\rho_T(0.648)|$  and  $|\tau_C(-0.022)| < |\tau\rho_T(0.467)|$ ; there is no significant relationship between X and Y at 5% significance level.

(ii) For Y and Z since  $|\rho_C(0.903)| > |\rho_T(0.648)|$  and  $|\tau_C(0.7826)| > |\tau\rho_T(0.467)|$ ; there is significant relationship between Y and Z at 5% significance level.

4. Three weighing scales from three different shops W, X and Y in a market were used to weigh 10 bags of beans (A, B, C.....) and the results in (kg) were given in the table below

	A	B	C	D	E	F	G	H	I	J
W	65	68	70	63	64	62	73	75	72	78
X	63	68	68	60	65	60	72	73	70	66
Y	63	74	78	75	64	73	79	70	67	79

Determine the rank correlation coefficient for the performance of the scales

(i) W and X ( $\rho = 0.8$ )

Spearman's rank correlation

W	X	RW	RX	d	d <sup>2</sup>
65	63	7	8	-1	1
68	68	6	4.5	1.5	2.25
70	68	5	4.5	0.5	0.25
63	60	9	9.5	-0.5	0.25
64	65	8	7	1	1
62	60	10	9.5	0.5	0.25
73	72	3	2	1	1
75	73	2	1	1	1
72	70	4	3	1	1
78	66	1	6	-5	25
					$\sum d^2 = 33$

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

$$= 1 - \left[ \frac{6 \times 33}{10(10^2-1)} \right] = 0.8$$

There is high positive relationship between Y and Z

Kendall's correlation coefficient

Rw	Rx	C	D
1	6	4	5
2	1	8	0
3	2	7	0
4	3	6	0
5	4.5	4	0
6	4.5	4	0
7	8	2	1
8	7	2	0
9	9.5	0	0
10	9.5	=37	=6

$$\tau = \frac{C-D}{C+D} = \frac{37-6}{37+6} = 0.721$$

There is high relationship positive between X and Y

- (ii) X and Y ( $\rho = 0.185$ )  
Spearman's correlation coefficient

X	Y	Rx	Ry	d	d <sup>2</sup>
63	63	8	10	-2	4
68	74	4.5	4	0.5	0.25
68	78	4.5	3	1.5	2.25
60	75	9.5	4	5.5	30.25
65	64	7	9	-2	4
60	73	9.5	6	3.5	12.25
72	79	2	1.5	0.5	0.25
73	70	1	7	-6	36
70	67	3	8	-5	25
66	79	6	1.5	4.5	20.25
					$\sum d^2 = 134.5$

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

$$= 1 - \left[ \frac{6 \times 134.5}{10(10^2 - 1)} \right] = 0.185$$

There is very low positive relationship between Y and Z

Kendall's rank coefficient

Rx	Ry	C	D
1	7	3	6
2	1.5	7	0
3	8	2	5
4.5	4	3	2
4.5	3	4	1
6	1.5	4	0
7	9	1	2
8	10	0	2
9.5	4	1	0
9.5	6	=25	=18

$$\tau = \frac{C-D}{C+D} = \frac{25-18}{25+18} = 0.163$$

There is very low relationship positive between X and Y

- (iii) Which of the three scales W, X and Y were in good working conditions  
W and X are in good working conditions because they show high positive correlation

5. (a) In many government institutions, officers complain about typing errors. A test was designed to investigate the relationship between typing speed and errors made. Twelve typist A, B, C ...L were picked at random to type a text. The table below shows the rankings of the typist according to speed and errors made. (N.B lowest ranking in error implies the least errors)

Typist	A	B	C	D	E	F	G	H	I	J	K	L
Speed	3	4	2	1	8	11	10	6	7	12	5	9
Errors	2	6	5	1	10	9	8	3	4	12	7	11

- (i) Calculate the coefficient of rank correlation Spearman's rank coefficient

Rs	Re	d	d <sup>2</sup>
3	2	1	1
4	6	-2	4
2	5	-3	9
1	1	0	0
8	10	-2	4
11	9	2	4
10	8	2	4
6	3	3	9
7	4	3	9
12	12	0	0
5	7	-2	4
9	11	-2	4
			$\sum d^2 = 52$

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

$$= 1 - \left[ \frac{6 \times 52}{12(12^2 - 1)} \right] = 0.8182$$

There is very low positive relationship between Y and Z

Kendall's rank correlation coefficient

Rs	Re	C	D
1	1	11	0
2	5	7	3
3	2	9	0
4	6	6	2
5	7	5	2
6	3	6	0
7	4	5	0
8	10	2	2
9	11	1	2
10	8	2	0
11	9	1	0
12	12	55	11

$$\tau = \frac{C-D}{C+D} = \frac{55-11}{55+11} = 0.667$$

There is very low relationship positive between X and Y

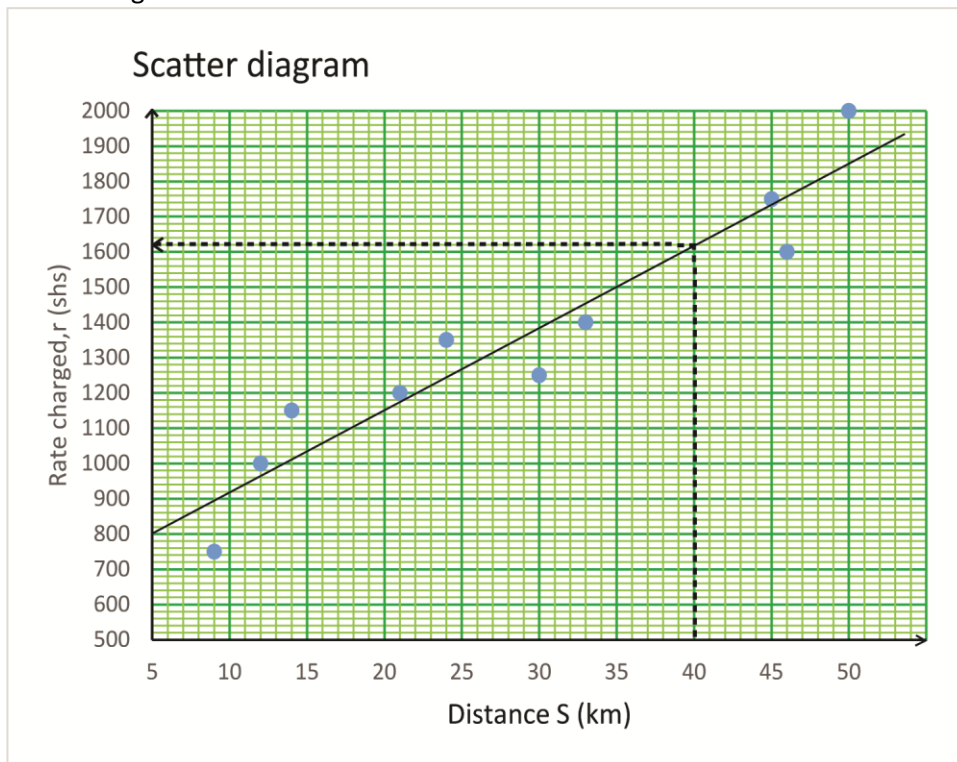
- (ii) Comment on the significance at 1% significance level  
(From the tables of critical values at 1% level of significance based on 12 observations,  $\rho = 0.727$ ,  $\tau = 0.545$ )

Since  $|\rho_C(0.8182)| > |\rho_T(0.727)|$  and  $|\tau_C(0.667)| > |\tau_T(0.545)|$ ; there is significant relationship between speed and errors at 1% significance level.

- (b) The cost of travelling at a certain distance away from the city centre is found to depend on the route and distance a given place is away from the centre. The table below gives average rates of travel charged for distances to be travelled away from the city centre

Distance (s km)	9	12	14	21	24	30	33	45	46	50
Rate charged (r shs)	750	1000	1150	1200	1350	1250	1400	1750	1600	2000

- (i) Plot the above data on a scatter diagram and draw a line of best fit through the points of the scatter diagram



- (ii) Estimate the expected value r corresponding to s = 40km  
Value of r corresponding to s = 40 is shs 1610

6. In a certain commercial institution, a speed and error typing examination was administered to 12 randomly selected candidates A, B, C, ..., L of the institution. The table below shows their speed (y) in seconds and the number of errors in their typed scripts (x)

	A	B	C	D	E	F	G	H	I	J	K	L
No. of errors (x)	12	24	20	10	32	30	28	15	18	40	27	35
Speed (y) in seconds	130	136	124	120	153	160	155	142	145	172	140	157

- (i) Calculate the coefficient of rank correlation of the ranking Spearman's rank correlation

Errors	speed	Re	Rs	d	d <sup>2</sup>
12	130	11	10	1	1
24	136	7	9	-2	4
20	124	8	11	-3	9
10	120	12	12	0	0
32	153	3	5	-2	4
30	160	4	2	2	4
28	155	5	4	1	1
15	142	10	7	3	9
18	145	9	6	3	9
40	172	1	1	0	0
27	140	6	8	-2	4
35	157	2	3	-1	1
					46

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

$$= 1 - \left[ \frac{6 \times 46}{12(12^2-1)} \right] = 0.839$$

There is high positive relationship between speed and errors made

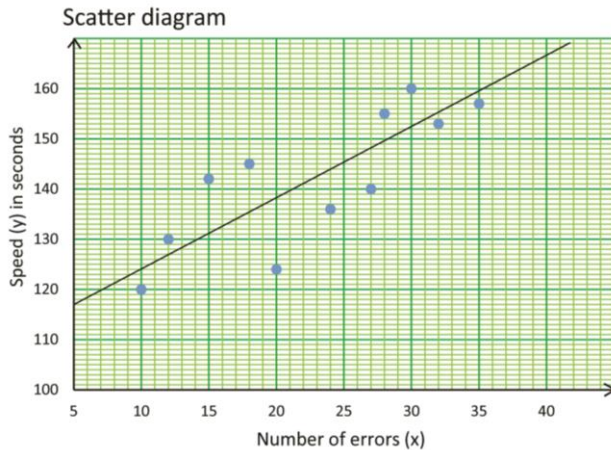
Kendall's rank correlation

Re	Rs	C	D
1	1	11	0
2	3	9	1
3	5	7	2
4	2	8	0
5	4	7	0
6	8	4	2
7	9	3	2
8	11	2	2
9	6	3	0
10	7	2	0
11	10	1	0
12	12	57	9

$$\tau = \frac{C-D}{C+D} = \frac{57-9}{57+9} = 0.72$$

There is high positive relationship between speed and errors made

- (ii) Comment on your results  
 (iii) Plot the above data on a scatter diagram and draw a line of the best fit through the points of the scatter diagram



7. The following table gives the marks obtained in calculus, physics and statistics by seven students

Calculus	72	50	60	55	35	48	82
Physics	61	55	70	50	30	50	73
Statistics	50	40	62	70	40	40	60

Determine the rank correlation coefficient for the performance of students in

(i) Calculus and physics

Spearman's correlation coefficient

C	P	Rc	Rp	d	d <sup>2</sup>
72	61	2	3	-1	1
50	55	5	4	1	1
60	70	3	2	1	1
55	50	4	5.5	-1.5	2.25
35	30	7	7	0	0
48	50	6	5.5	0.5	0.25
82	73	1	1	0	0
					$\sum d^2 = 5.5$

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

$$= 1 - \left[ \frac{6 \times 5.5}{7(7^2 - 1)} \right] = 0.902$$

There is high positive relationship between calculus and physics

Kendall's correlation coefficient

Rc	Rp	C	D
1	1	6	0
2	3	4	1
3	2	4	0
4	5.5	1	1
5	4	2	0
6	5.5	1	0
7	7	=18	=2

$$\tau = \frac{C-D}{C+D} = \frac{18-2}{18+2} = 0.8$$

There is high positive relationship between calculus and physics

- (ii) Calculus and statistics ( $\rho = 0.64$ )  
Spearman's correlation

C	S	Rc	Rs	d	d <sup>2</sup>
72	50	2	4	-2	4
50	40	5	6	-1	1
60	62	3	2	1	1
55	70	4	1	3	9
35	40	7	6	1	1
48	40	6	6	0	0
82	60	1	3	-2	4
					$\sum d^2 = 20$

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

$$= 1 - \left[ \frac{6 \times 20}{7(7^2-1)} \right] = 0.643$$

Kendall's rank correlation coefficient

Rc	Rs	C	D
1	3	4	2
2	4	3	2
3	2	3	1
4	1	3	0
5	6	0	0
6	6	0	0
7	6	=13	=5

$$\tau = \frac{C-D}{C+D} = \frac{13-5}{13+5} = 0.444$$

8. Given the table below

x	80	75	86	60	75	92	86	50	64	75
y	62	58	60	45	68	68	81	48	50	70

Determine the rank correlation coefficient between the variable x and y, comment on your results ( $\rho = 0.715$ )

Spearman's correlation

x	y	Rx	Ry	d	d <sup>2</sup>
80	62	4	5	-1	1
75	58	6	7	-1	1
86	60	2.5	6	-3.5	12.25
60	45	9	10	-1	1
75	68	6	3.5	2.5	6.25
92	68	1	3.5	-2.5	6.25
86	81	2.5	1	1.5	2.25
50	48	10	9	1	1
64	50	8	8	0	0
75	70	6	2	4	16
					47

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

$$= 1 - \left[ \frac{6 \times 47}{10(10^2-1)} \right] = 0.715$$

Kendall's correlation coefficient

Rx	Ry	C	D
1	3.5	6	2
2.5	6	4	4
2.5	1	7	0
4	5	4	2
6	7	3	2
6	3.5	3	2
6	2	3	0
8	8	2	0
9	10	0	1
10	9	=32	=13

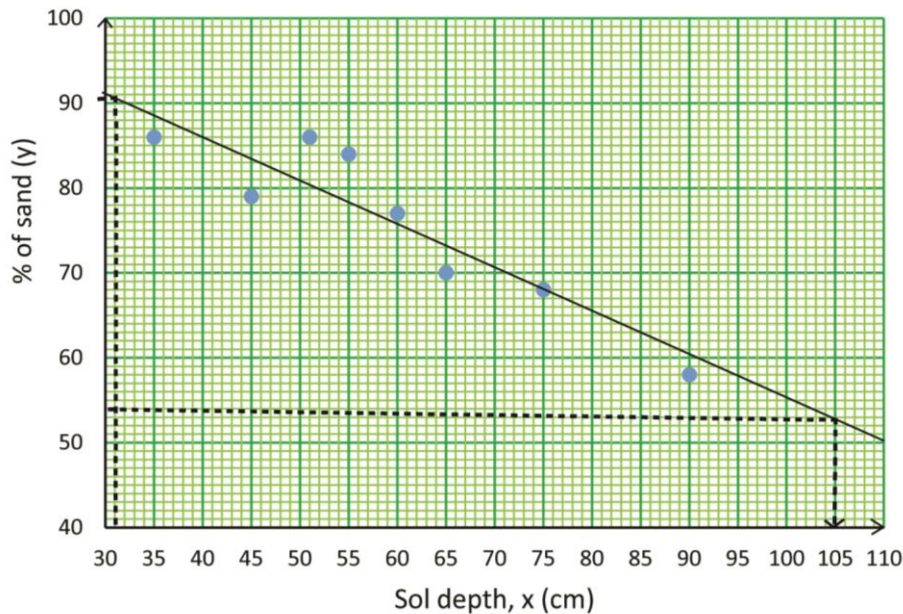
$$\tau = \frac{C-D}{C+D} = \frac{32-13}{32+13} = 0.422$$

9. The table below shows the percentage of sand  $y$  in the soils at different depth  $x$  (in cm)

Soil depth (x)(cm)	35	65	55	25	45	75	20	90	51	60
% of sand, y	86	70	84	92	79	68	96	58	86	77

(a) Plot the results on a scatter diagram. Comment on the relationship between the depth of the soil and the percentage of sand in the soil

Scatter diagram



(b) Draw the line of the best fit on you graph and use it to estimate

(i) The percentage of sand in the soil at a depth of 31cm (91%)

- (ii) Depth of the soil with 54% sand (105cm)
- (iii) Calculate the rank correlation coefficient  
Spearman's rank correlation coefficient

soil depth (x)	%of sand (y)	Rx	Ry	d	d <sup>2</sup>
35	86	8	3.5	4.5	20.25
65	70	3	8	-5	25
55	84	5	5	0	0
25	92	9	2	7	49
45	79	7	6	1	1
75	68	2	9	-7	49
20	96	10	1	9	81
90	58	1	10	-9	81
51	86	6	3.5	2.5	6.25
60	77	4	7	-3	9
					$\sum d^2 = 321.5$

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

$$= 1 - \left[ \frac{6 \times 321.5}{10(10^2 - 1)} \right] = -0.948$$

Kendall's correlation coefficient

Rx	Ry	C	D
1	10	0	9
2	9	0	8
3	8	0	7
4	7	0	6
5	5	1	4
6	3.5	1	2
7	6	0	3
8	3.5	0	2
9	2	0	1
10	1	2	42

$$\tau = \frac{C-D}{C+D} = \frac{2-42}{2+42} = -0.91$$

10. Eight applicants for a certain job obtained the following marks in aptitude and written test

Applicants	A	B	C	D	E	F	G	H
Aptitude test	33	45	16	42	45	35	40	48
Written test	57	60	40	75	68	48	54	68

(i) Calculate the coefficient of rank correlation of applicant's performance in the two tests

Aptitude (A)	Written (W)	R <sub>A</sub>	R <sub>W</sub>	d	d <sup>2</sup>
33	57	7	5	2	4
45	60	2	4	-2	4
16	40	8	8	0	0
42	75	4	1	3	9
45	68	3	2.5	0.5	0.25
35	48	6	7	-1	1
40	54	5	6	-1	1
48	68	1	2.5	-1.5	2.25
					21.5

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

$$= 1 - \left[ \frac{6 \times 21.5}{8(8^2-1)} \right] = 0.744$$

(ii) There is high positive relationship between Aptitude and written

Kendall's correlation coefficient

R <sub>x</sub>	R <sub>y</sub>	C	D
1	2.5	5	1
2	4	4	2
3	2.5	4	1
4	1	4	0
5	6	2	1
6	7	1	1
7	5	1	0
8	8	=21	=6

$$\tau = \frac{C-D}{C+D} = \frac{21-6}{21+6} = 0.555$$

(ii) There is moderate positive relationship between Aptitude and written

(ii) Comment on your results

11. The table below shows the marks scored by students in mathematics and fine art tests

students	A	B	C	D	E	F	G	H	I	J
Mathematics,	40	48	79	26	55	35	37	70	60	40
Fine art	59	62	68	47	46	39	63	29	55	67

Calculate the coefficient of rank correlation for the students' performance in the two subjects and comment on your results.

Math	Fine Art	R <sub>M</sub>	R <sub>F</sub>	d	d <sup>2</sup>
40	59	6.5	5	1.5	2.25
48	62	5	4	1	1
79	68	1	1	0	0
26	47	10	7	3	9
55	46	4	8	-4	16
35	39	9	9	0	0
37	63	8	3	5	25
70	29	2	10	-8	64
60	55	3	6	-3	9
40	67	6.5	2	4.5	20.25
					$\sum d^2 = 146.5$

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

$$= 1 - \left[ \frac{6 \times 146.5}{10(10^2 - 1)} \right]$$

$$= 0.112$$

Kendall's correlation coefficient

RM	RF	C	D
1	1	9	0
2	10	0	8
3	6	3	4
4	8	1	5
5	4	3	2
6.5	5	2	2
6.5	2	3	0
8	3	2	0
9	9	0	1
10	7	23	22

$$\tau = \frac{C-D}{C+D} = \frac{23-22}{23+22} = 0.022$$

12. Below are marks scored by 8 students A, B, C ... H in Mathematics, Economics and geography in the end of term examination.

	A	B	C	D	E	F	G	H
Math	52	75	41	60	81	31	65	52
Economics	50	60	35	65	66	45	69	48
Geography	35	40	60	54	63	40	65	72

Determine the rank correlation coefficient for the performance of students in

- (i) Math and economics

Spearman's correlation coefficient

Math, M	Economic C	R <sub>M</sub>	R <sub>E</sub>	d	d <sup>2</sup>
52	50	5.5	5	0.5	0.25
75	60	2	4	-2	4
41	35	7	8	-1	1
60	65	4	3	1	1
81	66	1	2	-1	1
31	45	8	7	1	1
65	69	3	1	2	4
52	48	5.5	6	-0.5	0.25
					∑ d <sup>2</sup> =12.5

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

$$= 1 - \left[ \frac{6 \times 12.5}{10(10^2-1)} \right]$$

$$= 0.851$$

Kendall's correlation coefficient

RM	RE	C	D
1	2	6	1
2	4	4	2
3	1	5	0
4	3	4	0
5.5	5	3	0
5.5	6	2	0
7	8	1	1
8	7	=25	=4

$$\tau = \frac{C-D}{C+D} = \frac{25-4}{25+4} = 0.724$$

- (ii) Geography and math  
Spearman's rank correlation coefficient

Geog, G	Math, M	R <sub>G</sub>	R <sub>M</sub>	d	d <sup>2</sup>
35	52	8	5.5	2.5	6.25
40	75	6.5	2	4.5	20.25
60	41	4	7	-3	9
54	60	5	4	1	1
63	81	3	1	2	4
40	31	6.5	8	-1.5	2.25
65	65	2	3	-1	1
72	52	1	5.5	-4.5	20.25
					64

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

$$= 1 - \left[ \frac{6 \times 64}{8(8^2-1)} \right]$$

$$= 0.238$$

Kendall's rank coefficient

RG	RM	C	D
1	5.5	2	4
2	3	4	2
3	1	5	0
4	7	1	3
5	4	2	1
6.5	2	2	0
6.5	8	0	1
8	5.5	16	11

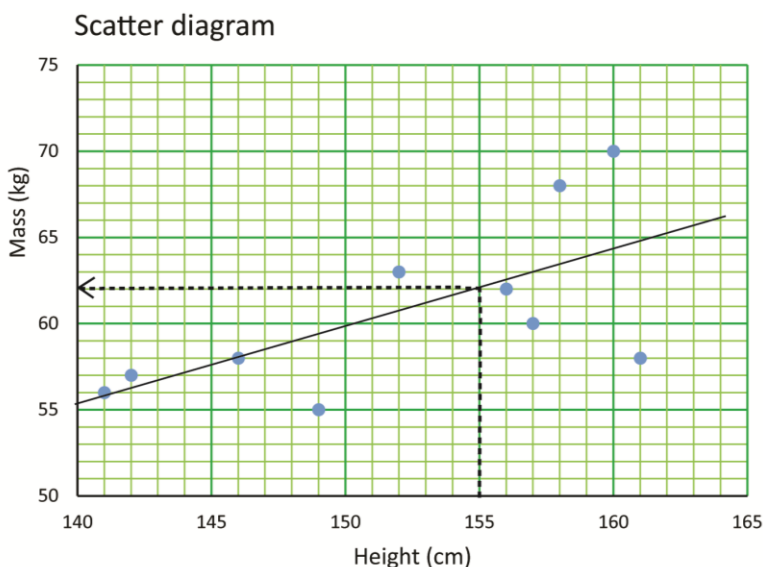
$$\tau = \frac{C-D}{C+D} = \frac{16-11}{16+11} = 0.185$$

- (iii) Comment on the significance of the math in performance of economics and geography. ( $\rho = 0.86$ ,  $\tau = 0.79$  based on 8 observations at a 1% level of significance)  
 Since  $|\rho_C(0.851)| < |\rho_T(0.86)|$  and  $|\tau_C(0.724)| < |\tau_T(0.79)|$ ; there is no significant relationship between math and economics at 1% significance level.

13. The heights and ages of ten students are given in the table below

Height, cm	156	151	152	146	160	157	149	142	158	140
Mass, kg	62	58	63	58	70	60	55	57	68	56

- (a) Plot the data on a scatter diagram



- (b) Draw the line of best fit on you graph and use it to estimate the mass corresponding to a height of 155cm(63kg)
- (c) Calculate the rank correlation coefficient for the data. Comment on the significance of the height on masses of students ( $\rho = 0.79$ ,  $\tau = 0.64$  based on 10 observations at 1% level of significance.)

**Method I: Using Spearman's rank correlation coefficient**

Height (x)	Mass (y)	R <sub>x</sub>	R <sub>y</sub>	R <sub>x</sub> - R <sub>y</sub> = d	d <sup>2</sup>
156	62	4	4	0	0
151	58	6	6.5	-0.5	0.25
152	63	5	3	2	4
146	58	8	6.5	1.5	2.25
160	70	1	1	0	0
157	60	3	5	-2	4
149	55	7	10	-3	9
142	57	9	8	1	1
158	68	2	2	0	0
140	56	10	9	1	1
					$\sum d^2 = 21.5$

$$\rho = 1 - \left[ \frac{6 \sum d^2}{n(n^2 - 1)} \right] = 1 - \left[ \frac{6 \times 21.5}{10(10^2 - 1)} \right] = 0.87$$

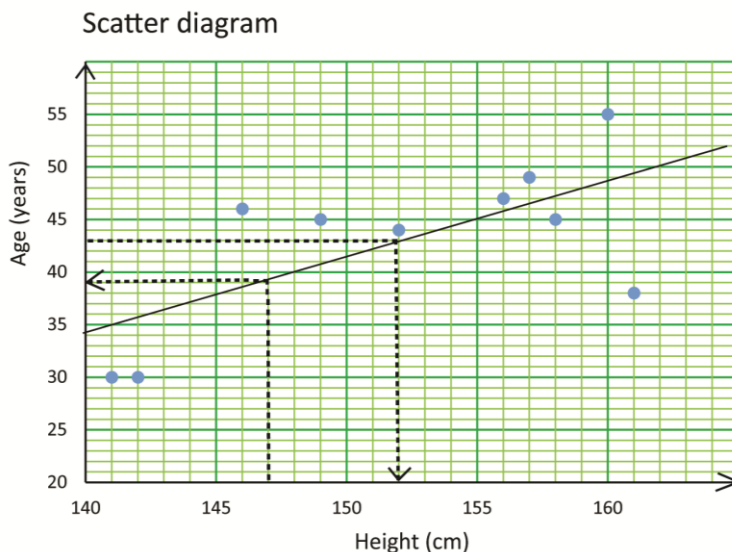
Since  $|\rho_C(0.87)| > |\rho_T(0.79)|$  there is no significant relationship between height and age at 1% significance level.

Since  $|\tau_C(0.6)| < |\tau\rho_T(0.64)|$ ;

14. The heights and ages of ten farmers are given in the table below

Height, cm	156	151	152	160	146	157	149	142	158	140
Age, years	47	38	44	55	46	49	45	30	45	20

(a) Plot the data on a scatter diagram



(b) Draw the line of best fit on your diagram and use it to estimate

- (i) Age when height = 147 (39)
- (ii) Height when the age is 43 (152)

(c) Calculate the rank correlation coefficient for the data. Comment on your results

**Method I: Using Spearman's rank correlation coefficient**

Height (x)	Age (y)	R <sub>x</sub>	R <sub>y</sub>	R <sub>x</sub> – R <sub>y</sub> = d	d <sup>2</sup>
156	47	4	3	1	1
151	38	6	8	-2	4
152	44	5	7	-2	4
160	55	1	1	0	0
146	46	8	4	4	16
157	49	3	2	1	1
149	45	7	5.5	1.5	2.25
142	30	9	9.5	-0.5	0.25
158	45	2	5.5	-3.5	12.25
140	30	10	9.5	0.5	0.25
					$\sum d^2 = 41$

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 41}{10(10^2-1)} = 0.7515(4D)$$

**Method II: using Kendall's rank correlation coefficient**

Let the farmers be A, B, C, D, E, F, G, H, I, J

Farmers	A	B	C	D	E	F	G	H	I	J
Height	156	151	152	160	146	157	149	142	158	140
Age	47	38	44	55	46	49	45	30	45	30

By re-arranging the findings we have

Farmers	D	I	F	A	C	B	G	E	H	J
Height	1	2	3	4	5	6	7	8	9	10
Age	1	5.5	3	3	7	8	5.5	4	9.5	9.5
agreements	9	4	7	6	3	2	2	2	0	=35
Disagreements	0	3	0	0	2	2	1	0	0	=8

s = total agreements – total disagreements

$$= 35 - 8 = 27$$

$$\tau = \frac{2s}{n(n-1)} = \frac{2 \times 27}{10(10-1)} = \frac{54}{90} = 0.6 \text{ Or } \tau = \frac{35-8}{35+8} = 0.63$$

**Comment**

Since  $|\rho_C(0.7515)| < |\rho_T(0.79)|$  and  $|\tau_C(0.6)| < |\tau_T(0.64)|$ ; there is no significant relationship between height and age at 1% significance level.

### Kendall's rank correlation coefficient

By naming the pairs we have

A(156, 62), B(151, 58), C(152, 63), D(146, 58), E(160, 70), F(157, 60), G(149, 55), H(142, 57), I(158, 68), J(141, 56)

	E	I	F	A	C	B	G	D	H	J
x	1	2	3	4	5	6	7	8	9	10
y	1	2	5	4	3	6.5	10	6.5	8	9
C	9	8	5	5	5	3	0	2	1	=38
D	0	0	2	1	0	0	3	0	0	=6

$$\tau(\text{tau}) = \frac{C-D}{C+D}$$

$$\tau(\text{tau}) = \frac{38-6}{38+6} = 0.73$$

Since  $\tau_c(0.73) > \tau_T(0.64)$ , a significant relationship exist between the heights and masses of student.

15. The table gives the points awarded to eight schools by three judges,  $J_1$ ,  $J_2$  and  $J_3$  during a music competition.  $J_1$  was the chief judge.

$J_1$	72	50	50	55	35	38	82	72
$J_2$	60	55	70	50	50	50	73	70
$J_3$	50	40	62	70	40	48	67	67

(a) Determine the rank correlation coefficients between the judges of

(i)  $J_1$  and  $J_2$

(ii)  $J_1$  and  $J_3$

(b) Who of the two judges had a better correlation with the chief judge? Give a reason.

Solution

(i)  $J_1$  and  $J_2$

$J_1$	$J_2$	$R_{J_1}$	$R_{J_2}$	$D_1$	$D_1^2$
72	60	2.5	4	-1.5	2.25
50	55	5.5	5	0.5	0.25
50	70	5.5	2.5	3	9
55	50	4	7	-3	9
35	50	8	7	1	1
38	50	7	7	0	0
82	73	1	1	0	0
72	70	2.5	2.5	0	0
					$\sum D_1^2 = 21.5$

$$\begin{aligned} \rho_1 &= 1 - \frac{6 \sum D_1^2}{n(n^2-1)} \\ &= 1 - \frac{6 \times 21.5}{8(8^2-1)} \\ &= \frac{125}{168} = 0.7440 \end{aligned}$$

(ii)  $J_1$  and  $J_3$  (10marks)

J <sub>1</sub>	J <sub>3</sub>	R <sub>J<sub>1</sub></sub>	R <sub>J<sub>3</sub></sub>	D <sub>2</sub>	D <sub>2</sub> <sup>2</sup>
72	50	2.5	5	-2.5	6.25
50	40	5.5	7.5	-4	4
50	62	5.5	4	0.5	2.25
55	70	4	1	3	9
35	40	8	7.5	0.5	0.25
38	48	7	6	1	1
82	67	1	2.5	-1.5	2.25
72	67	2.5	2.5	0	0
					∑ D <sub>1</sub> <sup>2</sup> = 25

$$\begin{aligned}
 \rho_2 &= 1 - \frac{6 \sum D_2^2}{n(n^2-1)} \\
 &= 1 - \frac{6 \times 25}{8(8^2-1)} \\
 &= 0.7023
 \end{aligned}$$

- (a) Who of the two other judges had a better correlation with the chief judge? Give a reason. (02marks)

J<sub>2</sub> has a better correlation with the Chief Judge because the coefficient of correlation is smaller showing a stronger mutual relationship

## Upper Critical Values for Kendall's Rank Correlation Coefficient $\hat{\tau}$

Note: In the table below, the critical values give significance levels as close as possible to but not exceeding the nominal  $\alpha$ .

$n$	Nominal $\alpha$					
	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	-	-	-	-
5	0.800	0.800	1.000	1.000	-	-
6	0.600	0.733	0.867	0.867	1.000	-
7	0.524	0.619	0.714	0.810	0.905	1.000
8	0.429	0.571	0.643	0.714	0.786	0.857
9	0.389	0.500	0.556	0.667	0.722	0.833
10	0.378	0.467	0.511	0.600	0.644	0.778
11	0.345	0.418	0.491	0.564	0.600	0.709
12	0.303	0.394	0.455	0.545	0.576	0.667
13	0.308	0.359	0.436	0.513	0.564	0.641
14	0.275	0.363	0.407	0.473	0.516	0.604
15	0.276	0.333	0.390	0.467	0.505	0.581
16	0.250	0.317	0.383	0.433	0.483	0.567
17	0.250	0.309	0.368	0.426	0.471	0.544
18	0.242	0.294	0.346	0.412	0.451	0.529
19	0.228	0.287	0.333	0.392	0.439	0.509
20	0.221	0.274	0.326	0.379	0.421	0.495
21	0.210	0.267	0.314	0.371	0.410	0.486
22	0.203	0.264	0.307	0.359	0.394	0.472
23	0.202	0.257	0.296	0.352	0.391	0.455
24	0.196	0.246	0.290	0.341	0.377	0.449
25	0.193	0.240	0.287	0.333	0.367	0.440
26	0.188	0.237	0.280	0.329	0.360	0.428
27	0.179	0.231	0.271	0.322	0.356	0.419
28	0.180	0.228	0.265	0.312	0.344	0.413
29	0.172	0.222	0.261	0.310	0.340	0.404

**Thank you Dr. Bbosa Science**