



Sponsored by
The Science Foundation College
Uganda East Africa
Senior one to senior six
+256 778 633 682, 753 802709
Based On, best for science

digitalteachers.co.ug



Regression and correlations

Correlations are the measure of how two or more variables are related to one another.

An independent variable is one which is not affected by the other (dependent variable) e.g. age is an independent variable while height of an individual is a dependent variable.

Regression is a statistical technique that relates a dependent variable to one or more independent variables. A regression model is able to show whether changes observed in the dependent variable are associated with changes in one or more of the independent variables by determining a best-fit line and seeing how the data is dispersed around this line.

Scatter diagrams/graphs

The scatter diagram /graphs are plots of pairs of numerical data, with one variable on each axis, to look for a relationship between them

Line of best fit

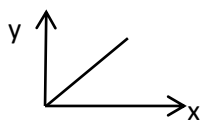
It is a straight line that goes roughly through the middle of all points plotted on a scatter diagram/graph.

It is used to express a relationship in a scatter plot of different data points.

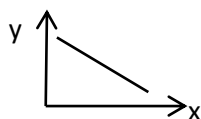
The closer the points to the line of the best fit, the stronger the correlation

Note that the line of the best fit passes through the mean of x and mean of y . these should be calculated to guide you as you draw the line of the best fit but should not be plotted.

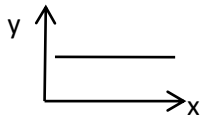
Positive correlation occurs when both the value of the variable increase or decrease



Negative or inverse correlation occurs when the increase in one of the variables cause a decrease in another.



Zero or correlation occurs when the increase or decrease in one of the variables does not cause any change in another.



Example 1

(a) The table below shows the marks of eight students in the mid-term test and end of term test in Economics.

Mid-term tests (x)	99	71	50	67	77	81	96	72
End of term test (y)	99	55	35	60	75	70	99	50

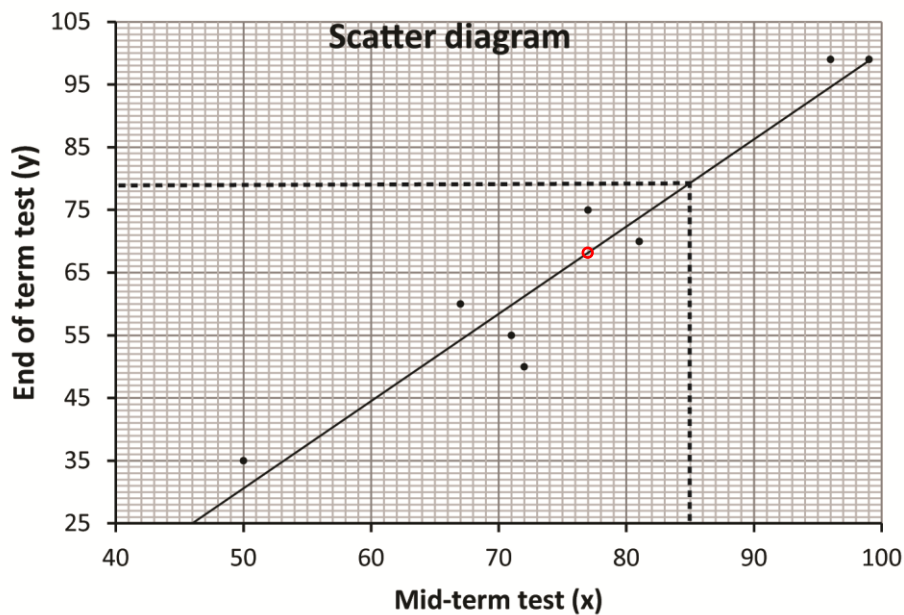
- (i) Draw a scatter diagram for the data
- (ii) On the same diagram draw a line of the best fit

Solution

Note that the line of the best fit passes through the mean of x and mean of y. these should be calculated to guide you as you draw the line of the best fit but should not be plotted. On the plot below (\bar{x}, \bar{y}) are encircled in red

$$\bar{x} = \frac{99+71+50+67+77+81+96+72}{8} = 76.6$$

$$\bar{y} = \frac{99+55+35+60+75+70+99+50}{8} = 67.9$$



- (iii) Use the line of the best fit to find the value of y when x – 85. (08 marks)
78
- (iv) Comment on the results
There is a positive correlation the two tests.

(b) The table shows scores by 10 students (A to J) in Physics and Mathematics tests

Students	A	B	C	D	E	F	G	H	I	J
Mathematics (x)	28	20	40	28	21	31	36	29	33	24
Physics (y)	30	20	40	28	22	35	35	27	31	23

(a) (i) Plot a scatter diagram for the given data.

(ii) Draw a line of the best fit on the scatter diagram

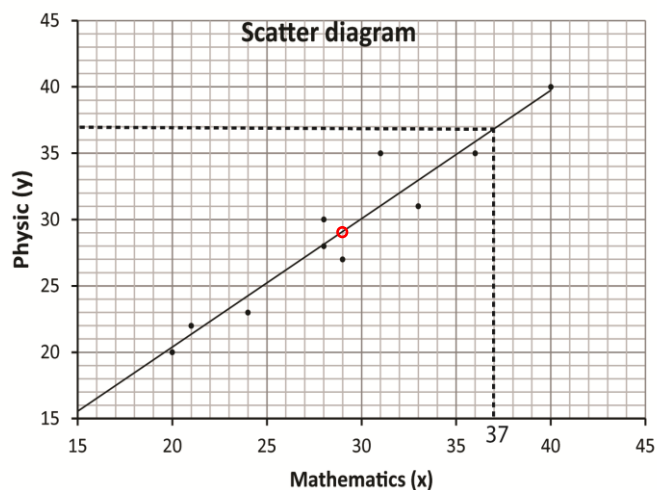
Solution

Note that the line of the best fit passes through the mean of x and the mean of y; these should be calculated to guide you when drawing a line of the best fit but should not be plotted

$$\bar{x} = \frac{28+20+40+28+21+31+36+29+33+24}{10} = 29$$

$$\bar{y} = \frac{30+20+40+28+22+35+35+27+31+23}{10} = 29.1$$

On the graph below the point (\bar{x}, \bar{y}) are encircled by red



(iii) Estimate the score in Mathematics for a student who scores 37 in physics. (08marks)

37

(c) The table below shows the heights to nearest cm and the masses to the nearest kg of 10 students, A to J.

Student	A	B	C	D	E	F	G	H	I	J
Mass (kg) (x)	53	68	57	52	66	64	63	58	57	68
Height (cm)(y)	148	172	156	139	163	158	168	151	144	170

- (a) (i) Plot the given data on a scatter diagram
(ii) Draw a line of best fit on the scatter diagram

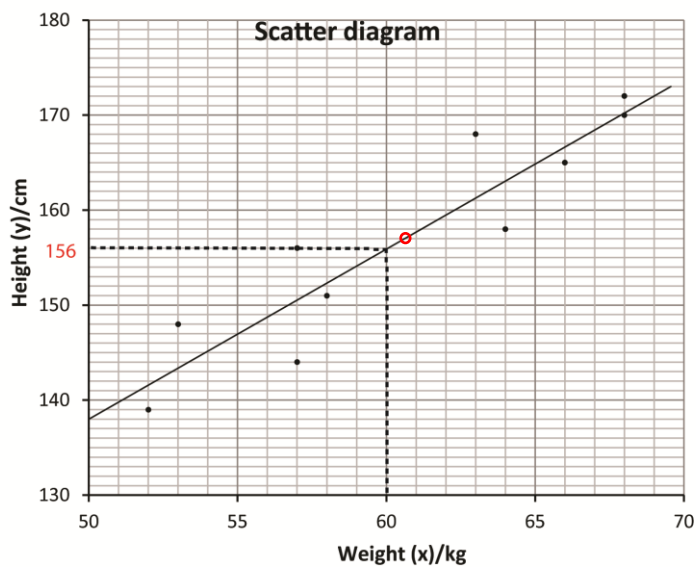
Solution

Note that the line of the best fit passes through the mean mass (\bar{x}) and mean of height (\bar{y}) and these should be calculated to guide you when drawing the line of the best fit

$$\bar{x} = \frac{53+68+57+52+66+64+63+58+57+68}{10} = 60.6$$

$$\bar{y} = \frac{148+172+156+139+163+158+168+151+144+170}{10} = 156.9$$

The point (\bar{x}, \bar{y}) is shown on the graph below by a red circle but should not be plotted



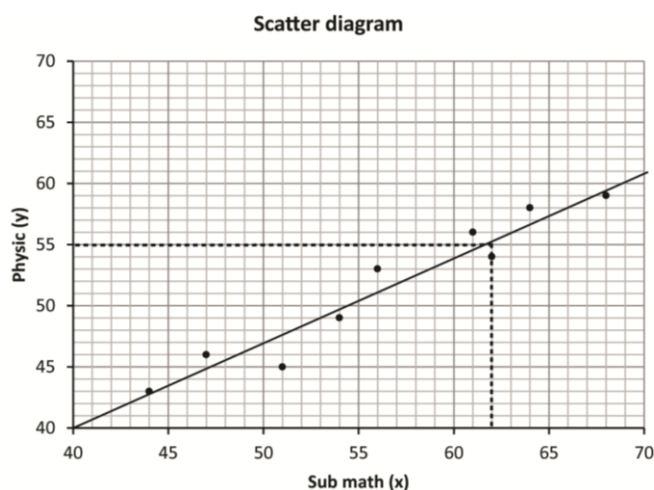
- (iii) Estimate the height of a student whose mass is 60 kg. (08marks)

156 (cm)

(d) The table below shows marks obtained in sub-Math and Physics by nine students.

Sub-Math (x)	51	62	64	47	54	44	68	61	56
Physics (y)	45	54	58	56	49	43	59	56	53

(a) (i) Draw a scatter diagram for the data



- (ii) On your scatter diagram, draw a line of best fit.
 (iii) Use the line of the best fit to estimate the value of x when y 55. (09 marks)
 62

Revision exercise 1

1. Given the information in the table below

X	10	15	20	15	30	35	40	45	50	60
y	15	20	35	40	35	50	55	40	54	56

- (a) Represent the above information on a scatter diagram
 (b) Draw a line of the best fit on your graph
 (c) Comment on your graph [there is a positive correlation between x and y]
2. The table below shows the marks obtained by 10 students in math (M) and physics (P) tests

M	80	75	65	90	95	98	78	65	54	60
P	70	85	70	90	92	88	76	70	73	76

- (a) Represent the above information on a scatter diagram
 (b) Draw a line of the best fit on your graph
 (c) Use your graph to estimate the math marks for a student that score 71% in physic. [60]

Spearman's Rank correlation

Spearman's rank correlation coefficient, denoted by ρ (rho), is a **nonparametric measure** of rank correlation. It assesses how well the relationship between two variables can be described using a monotonic function

Calculate Spearman's rank correlation, follow these steps:

1. **Convert the data to ranks:** Assign ranks to the values of each variable. If there are ties, assign the average rank.
2. **Compute the difference (d) in ranks:** For each pair of observations, calculate the difference between the ranks.
3. **Square the differences:** Square each difference.
4. **Sum the squared differences:** Add up all the squared differences.
5. **Apply the formula:** $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$ where d is the difference between the ranks of each observation, and (n) is the number of observations

Interpretations of rank correlation coefficient, ρ

Rank correlation coefficient, ρ	Interpretation
0	No (zero) correlation
$0 < \rho \leq 0.4$	Low positive correlation
$0.4 < \rho \leq 0.7$	Moderate positive correlation
$0.7 < \rho \leq 1$	high positive correlation

Note: The reverse is true for respective negative values

Example 1

- (a) The table below shows the marks of eight students in the mid-term test and end of term test in Economics.

Mid-term tests (x)	99	71	50	67	77	81	96	72
End of term test (y)	99	55	35	60	75	70	99	50

Calculate the Spearman's rank correlation coefficient. Comment on your results.

Solution

Mid-term tests (x)	End of term test (y)	R _x	R _y	d	d ²
99	99	1	1.5	-0.5	0.25
71	55	6	6	0	0
50	35	8	8	0	0
67	60	7	5	2	4
77	75	4	3	1	1
81	70	3	4	-1	1
96	99	2	1.5	0.5	0.25
72	50	5	7	-2	4
					$\sum d^2 = 10.5$

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 10.5}{8 \times 63} = 0.875$$

- (b) The table below shows the rank of marks awarded by Judge 1 (Rx) and Judge 2 (Ry) to 7 choir groups A to G.

Choir	A	B	C	D	E	F	G
Rank Judge (Rx)	2	4	6	1	5	3	7
Rank Judge (Ry)	2	3	5	1	6	4	7

Calculate Spearman's rank correlation coefficient between the marks awarded by the two judges.

Solution

Rearranging data

Rank Judge (Rx)	Rank Judge (Ry)	d	d ²
2	2	0	0
4	4	1	1
6	5	1	1
1	1	0	0
5	6	-1	1
3	4	-1	1
7	7	0	0
			$\Sigma d^2 = 4$

$$\rho = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 4}{7(49 - 1)} = 0.9286$$

Comment on your results.

There is a high positive correlation between the two Judges' marks

- (c) The table below shows the oral interview rank (X) and written interview rank (Y) for 12 candidates.

Candidate	A	B	C	D	E	F	G	H	I	J	K	L
Oral interview Rank (X)	8	10	9	4	12	5	11	7	3	6	1	2
Written interview Rank (Y)	11	12	9	7	10	6	8	5	2	4	1	3

Calculate Spearman's rank correlation coefficient and comment on your results.

Solution

Rearranging data

Candidate	R _x	R _y	d	d ²
A	8	11	-3	9
B	10	12	-2	4
C	9	9	0	0
D	4	7	3	9
E	12	10	2	4
F	5	6	-1	1
G	11	8	3	9
H	7	5	2	4
I	3	2	1	1
J	6	4	2	4
K	1	1	0	0
L	2	3	-1	1
				∑d ² = 46

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 46}{12(12^2-1)} = 0.892$$

Comment: there is high positive correlation between the oral interview and written interview.

(d) The table shows scores by 10 students (A to J) in Physics and Mathematics tests

Students	A	B	C	D	E	F	G	H	I	J
Mathematics (x)	28	20	40	28	21	31	36	29	33	24
Physics (y)	30	20	40	28	22	35	35	27	31	23

Calculate the rank correlation coefficient for the data and comment on your result.

Solution

Rearranging data

Student	R _x	R _y	d	d ²
A	6.5	5	2.5	2.25
B	10	10	0	0
C	1	1	0	0
D	6.5	6	0.5	0.25
E	9	9	0	0
F	4	2.5	1.5	2.25
G	2	2.5	-0.5	0.25
H	5	7	-2	4
I	3	4	-1	1
J	8	8	0	0
SUM				10

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 10}{10 \times 99} = 0.939$$

Comment: there is a high positive correlation between the marks of physics and mathematics

- (e) The table below shows the heights to nearest cm and the masses to the nearest kg of 10 students, A to J.

Student	A	B	C	D	E	F	G	H	I	J
Mass (kg) (x)	53	68	57	52	66	64	63	58	57	68
Height (cm)(y)	148	172	156	139	163	158	168	151	144	170

Calculate the rank correlation coefficient for the data.

Rearranging data

Student	R _x	R _y	d	d ²
A	10	8	2	4
B	1.5	1	0.5	0.25
C	7.5	6	1.5	1.25
D	9	10	0.5	0.25
E	3	4	-1	1
F	4	5	-1	1
G	5	3	2	4
H	6	7	-1	1
I	7.5	9	-1.5	2.25
J	1.5	2	-0.5	0.25
sum				17

$$\begin{aligned} \rho &= 1 - \frac{6 \sum d^2}{m(n^2 - 1)} \\ &= 1 - \frac{6 \times 17}{10(10^2 - 1)} \\ &= 0.897 \end{aligned}$$

Comment on your answer (07marks)

There is high positive correlation between height and weight of student.

Example 2

- (a) The table below shows marks obtained in sub-Math and Physics by nine students.

Sub-Math (x)	51	62	64	47	54	44	68	61	56
Physics (y)	45	54	58	56	49	43	59	56	53

Calculate the spearman's rank correlation coefficient and comment on the result. [$\rho = 0.967$]

- (b) The table below shows the marks obtained by 8 students in English and history tests

Eng. (E)	60	80	75	85	68	90	95	78
Hist. (H)	70	75	80	78	85	90	96	83

Calculate the spearman's rank correlation coefficient and comment on the result. [$\rho = 0.57$]

- (c) The table below shows the marks obtained by 10 students in Math (M) and Physics (P) tests

M	15	20	54	36	40	35	16	36	18	40
P	21	16	40	35	16	20	13	20	30	25

Calculate the spearman's rank correlation coefficient and comment on the result. [$\rho = 0.38$]

Please obtain free downloadable notes, exams and marking guides of Sub-math, ICT, general paper, biology, economics, geography etc. from digitalteachers.co.ug website

Thanks

Dr. Bbosa Science