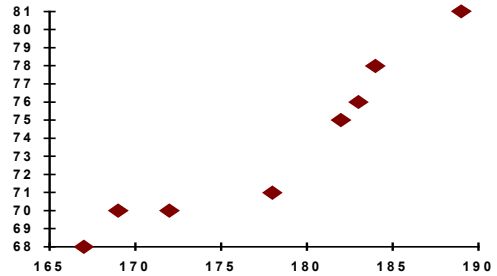
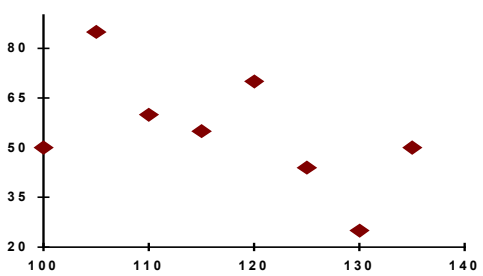
CORRELATION

1) From the following list of numbers, match each scatter graph with the most appropriate correlation coefficient. {–0.5, 0, –1, 0.24, –0.1, 1, 0.9}

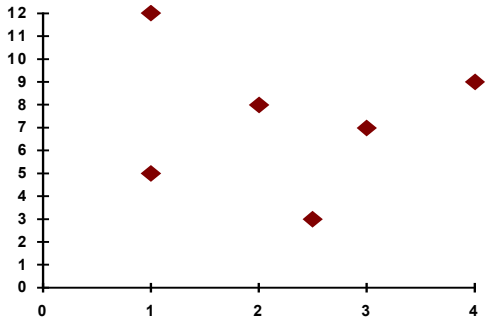
A



B



C



2) The following data relate to the percentage unemployment and percentage change in wages over several years.

%Unemployment, x	1.6	2.2	2.3	1.7	1.6	2.1	2.6	1.7	1.5	1.6
%Change in wages, y	5.0	3.2	2.7	2.1	4.1	2.7	2.9	4.6	3.5	4.4

a) Calculate the product-moment correlation coefficient between x and y.

It has been suggested that low unemployment and a low rate of wage inflation cannot exist together.

b) Without further calculation use your correlation coefficient to explain briefly whether or not you think the suggestion is justified.

3) A textile firm manufactures a particular fabric in a variety of different qualities at correspondingly different prices. In 1970, various amounts of the different qualities were sold; the amounts (in thousands of metres) and prices (in pence per metre) are shown below.

Amount sold	75	62	71	61	70	59	65	69
Price	37	55	25	20	47	62	24	52

Evaluate the product-moment correlation coefficient.

Draw a scatter diagram of the data, and contrast the information this provides about the relationship between prices and sales with that given by the product-moment correlation coefficient.

4) The mean temperature and the amount of milk sold to manufacturers during the months of a certain year were:

Month	Mean temperature	Milk sold (10 million gal.)
April	9	11
May	13	15
June	14	13
July	17	10
August	16	9
September	15	8

i) Calculate the product-moment correlation coefficient.

ii) Rank both sets of data (1 for highest, 2 for next highest etc.) and calculate the product-moment correlation coefficient between the two sets of ranks.

iii) Comment very briefly upon the two correlation coefficients.

5) Define the product-moment correlation coefficient and explain how you would interpret values of 0 and 1.

Guess the value of the correlation coefficient between the variables in the following situations and interpret/justify your estimate.

i) The marks in paper I and the total marks in a two-paper examination.

ii) The length of a child's foot and his/her I.Q.

iii) The height of water and the volume of road traffic at London Bridge, if high tide is at 7 a.m. The interval between successive high tides is about 12 hours.

6) The body and heart masses of fourteen 10-month old male mice are tabulated below.

Body mass (x) (g)	27	30	37	38	32	36	32	32	38	42	36	44	33	38
Heart mass (y) (mg)	118	136	156	150	140	155	157	114	144	159	149	170	131	160

a) Draw a scatter diagram of these data and draw a line of best fit which passes through the point with co-ordinates; $(\bar{x}, \bar{y})$.

b) Calculate the product-moment correlation coefficient for the data and comment briefly upon the closeness with which your line matches the data in the light of your calculation.
{Could you improve your line greatly?}

7) Draw a scatter graph of the following data.

x	1	2·5	6	7	4·5	3	6·5
y	0·5	1	3·5	6·5	3	2·5	5·5

Rank the above data and calculate the value of r_S.
Comment upon your value of r_S in the light of your scatter graph.

8) The table shows the marks awarded to six children in a competition. Calculate a coefficient of rank correlation for the data.

Child	A	B	C	D	E	F
Judge 1	6·8	7·3	8·1	9·8	7·1	9·2
Judge 2	7·8	9·4	7·9	9·6	8·9	6·9

9) i) State one **advantage** of Spearman's coefficient of rank correlation over the product-moment correlation coefficient.

ii) a) The marks of ten candidates in Physics and Mathematics are shown in the table.

Physics (x)	30	36	45	52	55	68	75	77	80	86
Maths (y)	45	60	55	56	70	65	72	77	85	80

Rank the results and hence calculate Spearman's coefficient of rank correlation.

b) For the above data,
$\sum x = 604$, $\sum y = 665$, $\sum xy = 42176$, $\sum x^2 = 39924$, and $\sum y^2 = 45649$.
Calculate the product-moment correlation coefficient, r, and compare your value with r_S found in a).

iii) State one **disadvantage** of Spearman's coefficient of rank correlation over the product-moment correlation coefficient.

10) The table shows the original marks of six candidates in two examinations.

Candidate	A	B	C	D	E	F
English	38	62	56	42	59	48
History	64	84	85	60	73	69

Calculate a coefficient of rank correlation and comment upon the value of your result.

The History papers are re-marked and one of the six candidates is awarded five additional marks. Given that the other marks, including the highest mark, and the coefficient of rank correlation are unchanged, state, with reasons, which candidate received the extra marks.

ANSWERS.

1) A : 0.9, B : –0.5, C : –0.1.

2) a) –0.558. b) Suggestion is not entirely justified since the correlation coefficient is not very significant.

3) $r = -0.218$. Scatter graph indicates no correlation.

4) i) $r = -0.380$.
ii) $r = -0.600$.
iii) The first coefficient is 'more sensitive' than the first since it uses the actual data. This is therefore likely to give a better indication of the true extent of any relationship.

5) The product moment correlation coefficient is a numerical measure of correlation between 2 variables x and y. It is given by the formula $r = \dfrac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}\sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$.

A correlation coefficient of 0 would imply no correlation whatsoever whilst a coefficient of 1 implies perfect positive correlation.

i) 0.8; The first variable forms a 'part' of the second variable and is thus very likely to be highly correlated.
ii) Any correlation would be purely spurious and nonsensical.
iii) 0.6 maybe. Each quantity is affected by a third variable; namely time. This would be another example of spurious correlation.

6) b) $r = 0.787$. Strong positive correlation. This means that there is little scope for improving the line's fit to the data.

7) $r_S = 1$.
$r_S = 1$ does NOT imply perfect positive correlation between the original data, only between the ranks! {This is because the rank-correlation coefficient yields less information than the stronger product-moment correlation coefficient. It is, however, easier to calculate!}

8) $r_S = 0{\cdot}26$.

9) i) r_S is easier to calculate than r.
ii) a) $r_S = 0{\cdot}939$ (3 significant figures).
b) $r = 0{\cdot}907$ (3 significant figures). Both r_S and r suggest a strong positive correlation between the test marks.
iii) r_S is not calculated from the original data and thus may not be as reliable a measure as the value of r.

10) $r_S = 0{\cdot}771$ (3 significant figures). This suggests a good level of positive correlation between the different subject marks.
Candidate E must have received the extra marks as adding 5 marks to any of the other scores will upset either the rankings (and thus the value of r_S) or the highest mark.