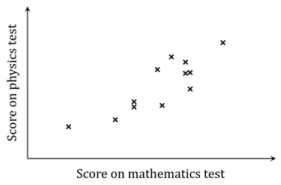
Question 1

A teacher collected the maths and physics test scores of a number of students and drew a scatter diagram to represent this data.

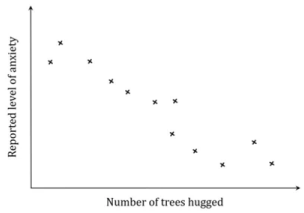


(a) Describe the correlation shown by the scatter diagram, and interpret the correlation in context.

[2]

a) (Fairly strong) positive correlation. The better a student performs on the maths test the better they tend to perform on the physics test.

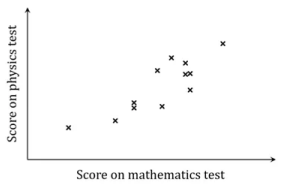
An alternative therapist collected data on his clients' reported levels of anxiety as well as the number of trees they had hugged in the course of therapy. He drew a scatter diagram to represent this data.



(b) Describe the correlation shown by the scatter diagram, and interpret the correlation in context.

[2]

A teacher collected the maths and physics test scores of a number of students and drew a scatter diagram to represent this data.

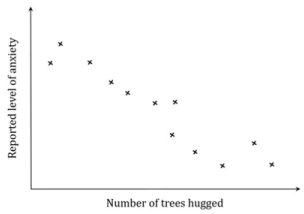


(a) Describe the correlation shown by the scatter diagram, and interpret the correlation in context.

[2]

b) (Strong) negative correlation. The more trees a client hugged the lower their reported level of anxiety.

An alternative therapist collected data on his clients' reported levels of anxiety as well as the number of trees they had hugged in the course of therapy. He drew a scatter diagram to represent this data.



(b) Describe the correlation shown by the scatter diagram, and interpret the correlation in context.

[2]

Question 2

Jennifer sells cups of tea at her shop and has noticed that she sells more tea on cooler days. On five different days, she records the maximum daily temperature, T, measured in degrees Celsius, and the number of cups of teas sold, C. The results are shown in the following table.

Maximum Daily Temperature, T.	3	5	8	9	12
Cups of tea sold, C.	37	34	33	26	21

(a) (i) Write down the equation of the regression line of C on T.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Use your regression equation from part (a)(i) to estimate the number of teas that Jennifer will sell on a day when the maximum temperature is 11°C.

[2]

(c) Being sure to consider the result from part (a)(ii) in your answer, state how confident you would be in your estimate from part (b).

[2]

a) Input data into your GDC and perform a linear regression $(ax + b)$.

x list: T

y list: C

i) $a = -1.756...$ $b = 43.195...$

$= -1.76$ (3sf) $= 43.2$ (3sf)

$C = -1.76T + 43.2$

ii) $r = -0.9425...$

$r = -0.942$ (3sf)

Jennifer sells cups of tea at her shop and has noticed that she sells more tea on cooler days. On five different days, she records the maximum daily temperature, T, measured in degrees Celsius, and the number of cups of teas sold, C. The results are shown in the following table.

Maximum Daily Temperature, T.	3	5	8	9	12
Cups of tea sold, C.	37	34	33	26	21

(a) (i) Write down the equation of the regression line of C on T.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

$C = -1.76T + 43.2$

[4]

(b) Use your regression equation from part (a)(i) to estimate the number of teas that Jennifer will sell on a day when the maximum temperature is 11°C.

[2]

(c) Being sure to consider the result from part (a)(ii) in your answer, state how confident you would be in your estimate from part (b).

[2]

b) Sub $T = 11$ into C

$C = -1.76(11) + 43.2$

$= 23.8780... \approx 24$

24 cups of tea

N.B calculator values for a and b used.

Jennifer sells cups of tea at her shop and has noticed that she sells more tea on cooler days. On five different days, she records the maximum daily temperature, T, measured in degrees Celsius, and the number of cups of teas sold, C. The results are shown in the following table.

Maximum Daily Temperature, T.	3	5	8	9	12
Cups of tea sold, C.	37	34	33	26	21

(a) (i) Write down the equation of the regression line of C on T.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Use your regression equation from part (a)(i) to estimate the number of teas that Jennifer will sell on a day when the maximum temperature is 11°C.

[2]

(c) Being sure to consider the result from part (a)(ii) in your answer, state how confident you would be in your estimate from part (b).

[2]

c) The estimate from part (b) is made by interpolation and the correlation is strong (r is close to -1).

∴ Very confident that the estimate is accurate.

Question 3

The following table shows the mean height, y cm, of primary school children who are age x years old.

Age, x years	6.25	7.35	8.5	9.25	10.75
Mean Height, y cm	115	121	129	136	140

(a) (i) Write down the equation of the regression line of y on x.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Use your regression equation from part (a)(i) to estimate the height of a child aged 9 years old.

[2]

(c) Explain why it is not appropriate to use the regression equation to estimate the age of a child who is 133 cm tall.

[1]

a) Input data into your GDC and perform a linear regression ($ax + b$).

x list: age

y list: height

i) $a = 5.8757... = 5.88$ (3sf) $\quad b = 78.7259... = 78.7$ (3sf)

$y = 5.88x + 78.7$

ii) $r = 0.9843...$

$r = 0.984$ (3sf)

The following table shows the mean height, y cm, of primary school children who are age x years old.

Age, x years	6.25	7.35	8.5	9.25	10.75
Mean Height, y cm	115	121	129	136	140

(a) (i) Write down the equation of the regression line of y on x.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

$y = 5.88x + 78.7$

[4]

(b) Use your regression equation from part (a)(i) to estimate the height of a child aged 9 years old.

[2]

(c) Explain why it is not appropriate to use the regression equation to estimate the age of a child who is 133 cm tall.

[1]

b) Sub $x = 9$ into y.

$y = 5.88(9) + 78.7$

$y = 131.6079...$

$y = 132$ cm

N.B calculator values for a and b used.

The following table shows the mean height, y cm, of primary school children who are age x years old.

Age, x years	6.25	7.35	8.5	9.25	10.75
Mean Height, y cm	115	121	129	136	140

(a) (i) Write down the equation of the regression line of y on x.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Use your regression equation from part (a)(i) to estimate the height of a child aged 9 years old.

[2]

(c) Explain why it is not appropriate to use the regression equation to estimate the age of a child who is 133 cm tall.

[1]

c) The regression line y on x should only be used to find y when given a value x.

Question 4

Rebecca, a regular jogger, ran the "Thao Dien Loop" on 7 consecutive days. The following table shows the distance, x km, that she ran and the corresponding number of calories, y, that she was able to burn during the run.

Distance (x)	2	5	6	7	10	12	14
Calories (y)	180	315	365	435	619	830	871

The number of calories burnt during a run is dependent upon on the length of the run.

(a) (i) Write down the equation of the regression line of y on x, giving your answer in the form $y = ax + b$ where a and b are constants to be found.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

On the 8th day, Rebecca is only able to run for 8 kilometres.

(c) Use the result from part (a)(i) to estimate the number of calories Rebecca will lose.

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

a) Input data into your GDC and perform a linear regression ($ax + b$).

x list: distance

y list: calories

i) $a = 62.2075...$ $= 62.2$ (3sf)

$b = 18.7681...$ $= 18.8$ (3sf)

$y = 62.2x + 18.8$

ii) $r = 0.9907...$

$r = 0.991$ (3sf)

b) Rebecca will burn an extra 62.2 calories for every extra 1 km ran.

Rebecca, a regular jogger, ran the "Thao Dien Loop" on 7 consecutive days. The following table shows the distance, x km, that she ran and the corresponding number of calories, y, that she was able to burn during the run.

Distance (x)	2	5	6	7	10	12	14
Calories (y)	180	315	365	435	619	830	871

The number of calories burnt during a run is dependent upon on the length of the run.

(a) (i) Write down the equation of the regression line of y on x, giving your answer in the form $y = ax + b$ where a and b are constants to be found.

$a = 62.2$

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

On the 8th day, Rebecca is only able to run for 8 kilometres.

(c) Use the result from part (a)(i) to estimate the number of calories Rebecca will lose.

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

Rebecca, a regular jogger, ran the "Thao Dien Loop" on 7 consecutive days. The following table shows the distance, x km, that she ran and the corresponding number of calories, y, that she was able to burn during the run.

Distance (x)	2	5	6	7	10	12	14
Calories (y)	180	315	365	435	619	830	871

The number of calories burnt during a run is dependent upon on the length of the run.

(a) (i) Write down the equation of the regression line of y on x, giving your answer in the form $y = ax + b$ where a and b are constants to be found.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

$y = 62.2x + 18.8$ [4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

On the 8th day, Rebecca is only able to run for 8 kilometres.

(c) Use the result from part (a)(i) to estimate the number of calories Rebecca will lose.

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

c) Sub $x = 8$ into y

$y = 62.2(8) + 18.8$

$y = 516.4285...$

$y = 516$ calories (3sf)

N.B calculator values for a and b used.

d) The answer from part (c) is valid and reliable as it was drawn by interpolation and r is very strong (close to 1).

Rebecca, a regular jogger, ran the "Thao Dien Loop" on 7 consecutive days. The following table shows the distance, x km, that she ran and the corresponding number of calories, y, that she was able to burn during the run.

Distance (x)	2	5	6	7	10	12	14
Calories (y)	180	315	365	435	619	830	871

The number of calories burnt during a run is dependent upon on the length of the run.

(a) (i) Write down the equation of the regression line of y on x, giving your answer in the form $y = ax + b$ where a and b are constants to be found.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

$y = 62.2x + 18.8$ [4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

On the 8th day, Rebecca is only able to run for 8 kilometres.

(c) Use the result from part (a)(i) to estimate the number of calories Rebecca will lose.

$y = 516$ calories (3sf) [2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

Question 5

The percentage of people who are willing to get a particular vaccine is dependent on their age. The following table shows the age, A years old, and the corresponding percentage of people, V, that are willing to receive a vaccine for 6 different ages.

Age, (A)	25	30	35	40	45	50
Percentage of willing people, (V)	57	59	61	62	68	75

(a) (i) Write down the equation of the regression line of V on A, giving your answer in the form $V = aA + b$ where a and b are constants to be found.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

(c) Use the result from part (a)(i) to estimate the percentage of people aged 95 years old who are in willing to receive a vaccine.

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

a) Input data into your GDC and perform a linear regression ($ax + b$).

x list: age

y list: percentage of willing people

i) $a = 0.6742...$
$= 0.674$ (3sf)

$b = 38.3809...$
$= 38.4$ (3sf)

$V = 0.674A + 38.4$

ii) $r = 0.9437...$

$r = 0.944$ (3sf)

The percentage of people who are willing to get a particular vaccine is dependent on their age. The following table shows the age, A years old, and the corresponding percentage of people, V, that are willing to receive a vaccine for 6 different ages.

Age, (A)	25	30	35	40	45	50
Percentage of willing people, (V)	57	59	61	62	68	75

(a) (i) Write down the equation of the regression line of V on A, giving your answer in the form $V = aA + b$ where a and b are constants to be found.

$a = 0.674$

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

(c) Use the result from part (a)(i) to estimate the percentage of people aged 95 years old who are in willing to receive a vaccine.

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

b) As a person's age increases by 1 year, their age groups approval of the vaccine increases by 0.674%.

The percentage of people who are willing to get a particular vaccine is dependent on their age. The following table shows the age, A years old, and the corresponding percentage of people, V, that are willing to receive a vaccine for 6 different ages.

Age, (A)	25	30	35	40	45	50
Percentage of willing people, (V)	57	59	61	62	68	75

(a) (i) Write down the equation of the regression line of V on A, giving your answer in the form $V = aA + b$ where a and b are constants to be found.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

$V = 0.674A + 38.4$

[4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

(c) Use the result from part (a)(i) to estimate the percentage of people aged 95 years old who are in willing to receive a vaccine.

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

c) Sub $A = 95$ into V.

$V = 0.674(95) + 38.4$

$V = 102.4380...$

$V = 102\%$ (3sf)

N.B calculator values for a and b used.

The percentage of people who are willing to get a particular vaccine is dependent on their age. The following table shows the age, A years old, and the corresponding percentage of people, V, that are willing to receive a vaccine for 6 different ages.

Age, (A)	25	30	35	40	45	50
Percentage of willing people, (V)	57	59	61	62	68	75

(a) (i) Write down the equation of the regression line of V on A, giving your answer in the form $V = aA + b$ where a and b are constants to be found.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Interpret, in the context of the question, the value of a found in part (a)(i).

[1]

(c) Use the result from part (a)(i) to estimate the percentage of people aged 95 years old who are in willing to receive a vaccine.

$V = 102\%$ (3sf)

[2]

(d) Comment on the validity of using the result from part (a)(i) to answer part (c).

[1]

d) The answer in part (c) was drawn via extrapolation, hence it is unreliable. Additionally the percentage is over 100% which is not possible.

Question 6

The price, $\$P$, of an airline ticket is dependent on the distance, d km, between two cities. The table below shows the airfares in US dollars from Prague in the Czech Republic, to eight different destinations in Europe.

Distance (d)	885	340	835	330	1270	295	650	1930
Price (P)	99	50	90	45	119	42.5	59	139

The relationship between d and P can be modelled by the regression line of P on d with equation $P = ad + b$.

(a) (i) Write down the equation of the regression line of P on d.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Use the result from part (a) to estimate the price of an airline ticket for a flight from Prague to a destination that is 2635 km away.

[2]

Madlenka buys a ticket for a flight from Prague to Cairo, a distance of 2635 km. The airfare in US dollars is \$245.

(c) Compare this price to your result from part (b), suggesting possible reasons for any discrepancies.

[2]

a) Input data into your GDC and perform a linear regression ($ax + b$).

x list: distance

y list: price

i) $a = 0.06289...$ $= 0.0629$ (3sf)

$b = 29.0623...$ $= 29.1$ (3sf)

$P = 0.0629d + 29.1$

ii) $r = 0.9634...$

$r = 0.963$ (3sf)

The price, $\$P$, of an airline ticket is dependent on the distance, d km, between two cities. The table below shows the airfares in US dollars from Prague in the Czech Republic, to eight different destinations in Europe.

Distance (d)	885	340	835	330	1270	295	650	1930
Price (P)	99	50	90	45	119	42.5	59	139

The relationship between d and P can be modelled by the regression line of P on d with equation $P = ad + b$.

(a) (i) Write down the equation of the regression line of P on d.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

$P = 0.0629d + 29.1$

[4]

(b) Use the result from part (a) to estimate the price of an airline ticket for a flight from Prague to a destination that is 2635 km away.

[2]

Madlenka buys a ticket for a flight from Prague to Cairo, a distance of 2635 km. The airfare in US dollars is \$245.

(c) Compare this price to your result from part (b), suggesting possible reasons for any discrepancies.

[2]

b) Sub $d = 2635$ into P.

$P = 0.0629(2635) + 29.1$

$P = 194.7836...$

$P = 195$ US dollars (3sf)

N.B calculator values for a and b used.

The price, $\$P$, of an airline ticket is dependent on the distance, d km, between two cities. The table below shows the airfares in US dollars from Prague in the Czech Republic, to eight different destinations in Europe.

Distance (d)	885	340	835	330	1270	295	650	1930
Price (P)	99	50	90	45	119	42.5	59	139

The relationship between d and P can be modelled by the regression line of P on d with equation $P = ad + b$.

(a) (i) Write down the equation of the regression line of P on d.

(ii) Write down the value of the Pearson's product-moment correlation coefficient, r.

[4]

(b) Use the result from part (a) to estimate the price of an airline ticket for a flight from Prague to a destination that is 2635 km away.

P = 195 US dollars (3sf)

[2]

Madlenka buys a ticket for a flight from Prague to Cairo, a distance of 2635 km. The airfare in US dollars is $245.

(c) Compare this price to your result from part (b), suggesting possible reasons for any discrepancies.

[2]

c) This is significantly more than the answer in part (b).
The mathematical reason for this is that the answer in part (b) was drawn via extrapolation (2635 km > 1930 km).
An additional reason is the other locations are all in Europe where Cairo is in Africa.

Question 7

A snack company is trialling a new series of crisp flavours. The eight flavours have been assigned the letters A through H, and the company has conducted taste tests in which volunteers assign each flavour a score on a scale from 1 to 10 (where '1' indicates 'I hate it!', and '10' indicates 'This is my new favourite!'). The following table collates the scores assigned to the flavours by each of three volunteers – Idris, Jameel and Kevin.

Flavours	A	B	C	D	E	F	G	H
Idris' score	1	9	4	8	10	3	7	6
Jameel's score	6	4	2	10	7	9	3	8
Kevin's score	9	4	7	2	1	7	5	6

The company would like to find the Spearman's rank correlation coefficients for these taste testers' rankings.

(a) Complete the information in the following table.

Flavours	A	B	C	D	E	F	G	H
Idris' rank	1	7	3	6	8	2	5	4
Jameel's rank	4	3	1	8	5	7	2	6
Kevin's rank	8	3	6.5	2	1	6.5	4	5

[3]

(b) Find the value of the Spearman's rank correlation coefficient, r_s, for:

(i) Idris' and Jameel's ranks

(ii) Idris' and Kevin's ranks

(iii) Jameel's and Kevin's ranks.

[3]

(c) Comment, in the context of the question, on the results obtained for r_s in part (b)(i), (ii) and (iii).

[3]

It is noticed that a greasy fingerprint has caused Kevin's score of 9 for flavour A to be misread. The score actually assigned by Kevin to flavour A was 8.

(d) Explain, with a reason, whether this will change any of the values for the Spearman's rank correlation coefficient r_s that were calculated in part (b).

[2]

a) Rank the scores in ascending order.
ie. Score = 1 ∴ Rank = 1
Score = 10 ∴ Rank = 8

A snack company is trialling a new series of crisp flavours. The eight flavours have been assigned the letters A through H, and the company has conducted taste tests in which volunteers assign each flavour a score on a scale from 1 to 10 (where '1' indicates 'I hate it!', and '10' indicates 'This is my new favourite!'). The following table collates the scores assigned to the flavours by each of three volunteers – Idris, Jameel and Kevin.

Flavours	A	B	C	D	E	F	G	H
Idris' score	1	9	4	8	10	3	7	6
Jameel's score	6	4	2	10	7	9	3	8
Kevin's score	9	4	7	2	1	7	5	6

The company would like to find the Spearman's rank correlation coefficients for these taste testers' rankings.

(a) Complete the information in the following table.

Flavours	A	B	C	D	E	F	G	H
Idris' rank	1	7	3	6	8	2	5	4
Jameel's rank	4	3	1	8	5	7	2	6
Kevin's rank	8	3	6.5	2	1	6.5	4	5

[3]

(b) Find the value of the Spearman's rank correlation coefficient, r_s, for:

(i) Idris' and Jameel's ranks

(ii) Idris' and Kevin's ranks

(iii) Jameel's and Kevin's ranks.

[3]

(c) Comment, in the context of the question, on the results obtained for r_s in part (b)(i), (ii) and (iii).

[3]

It is noticed that a greasy fingerprint has caused Kevin's score of 9 for flavour A to be misread. The score actually assigned by Kevin to flavour A was 8.

(d) Explain, with a reason, whether this will change any of the values for the Spearman's rank correlation coefficient r_s that were calculated in part (b).

[2]

b) Input the ranks for each part into your GDC and perform a linear regression.

i) x list: Idris y list: Jameel

$r_s = 0.04761...$ $r_s = 0.0476$ (3sf)

ii) x list: Idris y list: Kevin

$r_s = -0.9707...$ $r_s = -0.971$ (3sf)

iii) x list: Jameel y list: Kevin

$r_s = -0.2395...$ $r_s = -0.240$ (3sf)

A snack company is trialling a new series of crisp flavours. The eight flavours have been assigned the letters A through H, and the company has conducted taste tests in which volunteers assign each flavour a score on a scale from 1 to 10 (where '1' indicates 'I hate it!', and '10' indicates 'This is my new favourite!'). The following table collates the scores assigned to the flavours by each of three volunteers – Idris, Jameel and Kevin.

Flavours	A	B	C	D	E	F	G	H
Idris' score	1	9	4	8	10	3	7	6
Jameel's score	6	4	2	10	7	9	3	8
Kevin's score	9	4	7	2	1	7	5	6

The company would like to find the Spearman's rank correlation coefficients for these taste testers' rankings.

(a) Complete the information in the following table.

Flavours	A	B	C	D	E	F	G	H
Idris' rank	1				8			
Jameel's rank			1	8				
Kevin's rank	8				1	6.5		

[3]

(b) Find the value of the Spearman's rank correlation coefficient, r_s, for:

(i) Idris' and Jameel's ranks $r_s = 0.0476$ (3sf)

(ii) Idris' and Kevin's ranks $r_s = -0.971$ (3sf)

(iii) Jameel's and Kevin's ranks. $r_s = -0.240$ (3sf)

[3]

(c) Comment, in the context of the question, on the results obtained for r_s in part (b)(i), (ii) and (iii).

[3]

It is noticed that a greasy fingerprint has caused Kevin's score of 9 for flavour A to be misread. The score actually assigned by Kevin to flavour A was 8.

(d) Explain, with a reason, whether this will change any of the values for the Spearman's rank correlation coefficient r_s that were calculated in part (b).

[2]

c) i) The correlation between Idris and Jameel is almost zero, so there is no way of guessing what flavours one likes based on what the other likes.

ii) Idris and Kevin have a strong negative correlation (close to -1), meaning Idris hates what Kevin likes and vice versa.

iii) Jameel and Kevin have a weak negative correlation, meaning there is a slight tendency for one to like what the other does not and vice versa.

A snack company is trialling a new series of crisp flavours. The eight flavours have been assigned the letters A through H, and the company has conducted taste tests in which volunteers assign each flavour a score on a scale from 1 to 10 (where '1' indicates 'I hate it!', and '10' indicates 'This is my new favourite!'). The following table collates the scores assigned to the flavours by each of three volunteers – Idris, Jameel and Kevin.

Flavours	A	B	C	D	E	F	G	H
Idris' score	1	9	4	8	10	3	7	6
Jameel's score	6	4	2	10	7	9	3	8
Kevin's score	8	4	7	2	1	7	5	6

The company would like to find the Spearman's rank correlation coefficients for these taste testers' rankings.

(a) Complete the information in the following table.

Flavours	A	B	C	D	E	F	G	H
Idris' rank	1	7	3	6	8	2	5	4
Jameel's rank	4	3	1	8	5	7	2	6
Kevin's rank	8	3	6.5	2	1	6.5	4	5

[3]

(b) Find the value of the Spearman's rank correlation coefficient, r_s, for:

(i) Idris' and Jameel's ranks

(ii) Idris' and Kevin's ranks

(iii) Jameel's and Kevin's ranks.

[3]

(c) Comment, in the context of the question, on the results obtained for r_s in part (b)(i), (ii) and (iii).

[3]

It is noticed that a greasy fingerprint has caused Kevin's score of 9 for flavour A to be misread. The score actually assigned by Kevin to flavour A was 8.

(d) Explain, with a reason, whether this will change any of the values for the Spearman's rank correlation coefficient r_s that were calculated in part (b).

[2]

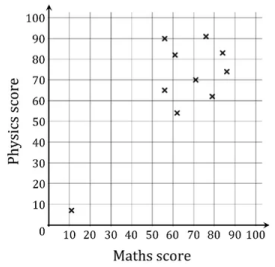
d) The rank of flavour A does not change, therefore this will not change any of the answers in part (b).

Question 8

The following table collates the scores achieved on a recent maths test and a recent physics test by a group of 10 students.

Student	A	B	C	D	E	F	G	H	I	J
Maths score	62	79	71	56	61	84	86	76	56	11
Physics score	54	62	70	65	82	83	74	91	90	7

The scatter diagram for these scores is shown in the diagram below:



(a) Write down the value of the Pearson's product-moment correlation coefficient, r,

(i) with student J included

(ii) without student J included.

[3]

(b) Complete the information in the following table, ranking the scores from highest to lowest:

Student	A	B	C	D	E	F	G	H	I	J
Maths rank			5	8.5			1			10
Physics rank							5	1		10

[2]

(c) Find the value of the Spearman's rank correlation coefficient, r_s,

(i) with student J included

(ii) without student J included.

[3]

(d) Comment on the results of parts (a) and (c).

[3]

a) Input data into your GDC and perform a linear regression.

i) With student J.

x list: maths y list: physics

$r = 0.7695...$ $r = 0.770$ (3sf)

ii) Input data into your GDC and perform a linear regression. Without student J.

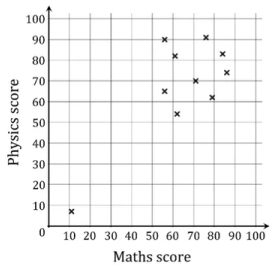
x list: maths y list: physics

$r = 0.08696...$ $r = 0.0870$ (3sf)

The following table collates the scores achieved on a recent maths test and a recent physics test by a group of 10 students.

Student	A	B	C	D	E	F	G	H	I	J
Maths score	62	79	71	56	61	84	86	76	56	11
Physics score	54	62	70	65	82	83	74	91	90	7

The scatter diagram for these scores is shown in the diagram below:



(a) Write down the value of the Pearson's product-moment correlation coefficient, r,

(i) with student J included

(ii) without student J included.

[3]

(b) Complete the information in the following table, ranking the scores from highest to lowest:

Student	A	B	C	D	E	F	G	H	I	J
Maths rank	6	3	5	8.5	7	2	1	4	8.5	10
Physics rank	9	8	6	7	4	3	5	1	2	10

[2]

(c) Find the value of the Spearman's rank correlation coefficient, r_s,

(i) with student J included

(ii) without student J included.

[3]

(d) Comment on the results of parts (a) and (c).

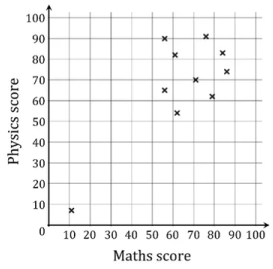
[3]

b) Order the scores from 1-10 and fill in the table.

The following table collates the scores achieved on a recent maths test and a recent physics test by a group of 10 students.

Student	A	B	C	D	E	F	G	H	I	J
Maths score	62	79	71	56	61	84	86	76	56	11
Physics score	54	62	70	65	82	83	74	91	90	7

The scatter diagram for these scores is shown in the diagram below:



(a) Write down the value of the Pearson's product-moment correlation coefficient, r,

(i) with student J included

(ii) without student J included.

[3]

(b) Complete the information in the following table, ranking the scores from highest to lowest:

Student	A	B	C	D	E	F	G	H	I	J
Maths rank	6	3	5	8.5	7	2	1	4	8.5	10
Physics rank	9	8	6	7	4	3	5	1	2	10

[2]

(c) Find the value of the Spearman's rank correlation coefficient, r_s,

(i) with student J included

(ii) without student J included.

[3]

(d) Comment on the results of parts (a) and (c).

[3]

c) Input the ranks for each part into your GDC and perform a linear regression.

i) With student J.

x list: maths y list: physics

$r_s = 0.3039...$ $r_s = 0.304$ (3sf)

ii) Without student J.

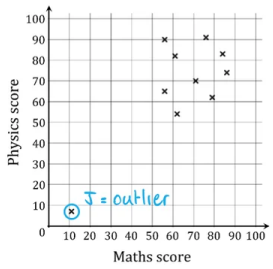
x list: maths y list: physics

$r_s = 0.04184...$ $r_s = 0.0418$ (3sf)

The following table collates the scores achieved on a recent maths test and a recent physics test by a group of 10 students.

Student	A	B	C	D	E	F	G	H	I	J
Maths score	62	79	71	56	61	84	86	76	56	11
Physics score	54	62	70	65	82	83	74	91	90	7

The scatter diagram for these scores is shown in the diagram below:



(a) Write down the value of the Pearson's product-moment correlation coefficient, r,

(i) with student J included $r = 0.770$ (3sf)

(ii) without student J included. $r = 0.0870$ (3sf)

[3]

(b) Complete the information in the following table, ranking the scores from highest to lowest:

Student	A	B	C	D	E	F	G	H	I	J
Maths rank	6	3	5	8.5	7	2	1	4	8.5	10
Physics rank	9	8	6	7	4	3	5	1	2	10

[2]

(c) Find the value of the Spearman's rank correlation coefficient, r_s,

(i) with student J included $r_s = 0.304$ (3sf)

(ii) without student J included. $r_s = 0.0418$ (3sf)

[3]

(d) Comment on the results of parts (a) and (c).

[3]

d) Student J is an outlier.
In part (a) we used PMCC (r) and in part (c) we used Spearman's rank c.c (r_s).

Without student J both versions of the c.c show virtually no correlation (close to 0).

With student J included both versions of the c.c are affected, however r_s is less affected.

This is because r_s is less affected by outliers.

Question 9

A school decides to introduce a test for students aged 14 as they start their Mathematics course to predict how well they will perform in their final Mathematics examination at age 16. 10 students are randomly selected to sit the test and their results are compared to their final exam score.

The results of the 10 students are shown in the table below.

Student	1	2	3	4	5	6	7	8	9	10
Test (%)	88	67	26	45	76	79	97	56	42	63
Final Exam (%)	75	68	31	51	72	66	94	25	38	58

(a) State the name for this test of validity.

[1]

a) Criterion-related

(b) Calculate Pearson's product moment correlation coefficient for this data.

[2]

(c) Hence determine if the test is a valid indicator of future performance in the formal Mathematics examinations. Give a reason for your answer.

[2]

A school decides to introduce a test for students aged 14 as they start their Mathematics course to predict how well they will perform in their final Mathematics examination at age 16. 10 students are randomly selected to sit the test and their results are compared to their final exam score.

The results of the 10 students are shown in the table below.

Student	1	2	3	4	5	6	7	8	9	10
Test (%)	88	67	26	45	76	79	97	56	42	63
Final Exam (%)	75	68	31	51	72	66	94	25	38	58

(a) State the name for this test of validity.

[1]

(b) Calculate Pearson's product moment correlation coefficient for this data.

[2]

(c) Hence determine if the test is a valid indicator of future performance in the formal Mathematics examinations. Give a reason for your answer.

[2]

b) Input the data into your GDC

$r = 0.8776... = 0.878$ (3s.f)

c) The test is valid as r is closer to 1 than 0.

A school decides to introduce a test for students aged 14 as they start their Mathematics course to predict how well they will perform in their final Mathematics examination at age 16. 10 students are randomly selected to sit the test and their results are compared to their final exam score.

The results of the 10 students are shown in the table below.

Student	1	2	3	4	5	6	7	8	9	10
Test (%)	88	67	26	45	76	79	97	56	42	63
Final Exam (%)	75	68	31	51	72	66	94	25	38	58

(a) State the name for this test of validity.

[1]

(b) Calculate Pearson's product moment correlation coefficient for this data.

[2]

(c) Hence determine if the test is a valid indicator of future performance in the formal Mathematics examinations. Give a reason for your answer.

[2]

Question 10

A company is considering 8 candidates for a role and asks them to complete a set of questions designed to identify potential leadership qualities at the first stage of interview. A few days later, they ask them to complete a second set of similarly designed questions in order to test the reliability of the questionnaire.

The results have been recorded in the table below.

Interviewee	A	B	C	D	E	F	G	H
Questionnaire 1	75	68	44	52	59	71	80	63
Questionnaire 2	61	70	57	60	53	78	67	41

(a) State the name of this test for reliability.

[1]

(b) State a possible disadvantage of this test for reliability.

[1]

(c) Find Pearson's product moment correlation coefficient for the data in the table above.

[2]

(d) Hence determine if the data from the personality test is reliable. Give a reason for your answer.

[2]

b) It can be difficult to create two sets of questions that are equal in difficulty OR it is time consuming to create two sets of questions.

c) Input the data into your GDC

$r = 0.4253\ldots = 0.425$ (3s.f)

A company is considering 8 candidates for a role and asks them to complete a set of questions designed to identify potential leadership qualities at the first stage of interview. A few days later, they ask them to complete a second set of similarly designed questions in order to test the reliability of the questionnaire.

The results have been recorded in the table below.

Interviewee	A	B	C	D	E	F	G	H
Questionnaire 1	75	68	44	52	59	71	80	63
Questionnaire 2	61	70	57	60	53	78	67	41

(a) State the name of this test for reliability.

[1]

(b) State a possible disadvantage of this test for reliability.

[1]

(c) Find Pearson's product moment correlation coefficient for the data in the table above.

[2]

(d) Hence determine if the data from the personality test is reliable. Give a reason for your answer.

[2]

A company is considering 8 candidates for a role and asks them to complete a set of questions designed to identify potential leadership qualities at the first stage of interview. A few days later, they ask them to complete a second set of similarly designed questions in order to test the reliability of the questionnaire.

The results have been recorded in the table below.

Interviewee	A	B	C	D	E	F	G	H
Questionnaire 1	75	68	44	52	59	71	80	63
Questionnaire 2	61	70	57	60	53	78	67	41

(a) State the name of this test for reliability.

[1]

(b) State a possible disadvantage of this test for reliability.

[1]

(c) Find Pearson's product moment correlation coefficient for the data in the table above.

[2]

(d) Hence determine if the data from the personality test is reliable. Give a reason for your answer.

[2]

d) The questionnaires are unreliable as r is closer to 0 than 1.