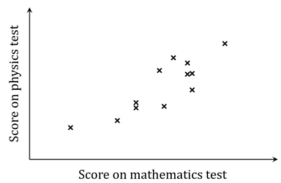
Question 1

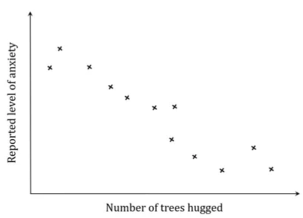
A teacher collected the maths and physics test scores of a number of students and drew a scatter diagram to represent this data.



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

[2]

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.

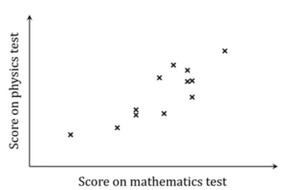


(b) Describe the correlation shown by the scatter diagram, and interpret the correlation in context.

[2]

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

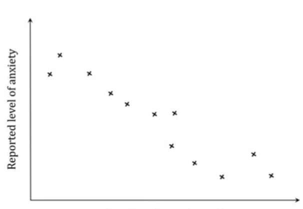
A teacher collected the maths and physics test scores of a number of students and drew a scatter diagram to represent this data.



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

[2]

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]

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

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

The relationship between T and C can be modelled by the regression line of C on T with equation $C = aT + b$.

(a) (i) Find the value of a and the value of b.

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

[4]

(b) Using the regression equation, 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) $a = -1.756...$ $\quad$ $b = 43.195...$

$\boxed{a = -1.76 \text{ (3sf)}}$ $\quad$ $\boxed{b = 43.2 \text{ (3sf)}}$

ii) $r = -0.9425...$

$\boxed{r = -0.942 \text{ (3sf)}}$

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

The relationship between T and C can be modelled by the regression line of C on T with equation $C = aT + b$.

(a) (i) Find the value of a and the value of b. $\boxed{a = -1.76}$ $\boxed{b = 43.2}$

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

[4]

(b) Using the regression equation, 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

$C = -1.76(11) + 43.2$

$= 23.8780... \approx 24$

$\boxed{\text{24 cups of tea}}$

NB 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

The relationship between T and C can be modelled by the regression line of C on T with equation $C = aT + b$.

(a) (i) Find the value of a and the value of b.

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

[4]

(b) Using the regression equation, 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).

$\therefore$ 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

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b.

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

[4]

(b) Using the regression equation with your values of a and b from part (a)(i), 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) $a = 5.8757...$ $b = 78.7259...$

$\boxed{a = 5.88 \ (3sf)}$ $\boxed{b = 78.7 \ (3sf)}$

ii) $r = 0.9843...$

$\boxed{r = 0.984 \ (3sf)}$

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

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b. $\boxed{a = 5.88}$ $\boxed{b = 78.7}$

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

[4]

(b) Using the regression equation with your values of a and b from part (a)(i), 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.

$y = 5.88(9) + 78.7$

$y = 131.6079\ldots$

$\boxed{y = 132\text{ cm } (3sf)}$

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

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b.

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

[4]

(b) Using the regression equation with your values of a and b from part (a)(i), 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 of 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. The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b.

(ii) Write down the value of 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) answer part (c).

[1]

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

x list: distance

y list: calories

i) $a = 62.2075...$ $b = 18.7681...$

$\boxed{a = 62.2 \quad (3sf)}$ $\boxed{b = 18.8 \quad (3sf)}$

ii) $r = 0.9907...$

$\boxed{r = 0.991 \quad (3sf)}$

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. The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

$\boxed{a = 62.2}$

(a) (i) Find the value of a and the value of b.

(ii) Write down the value of 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) answer part (c).

[1]

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

EXAM PAPERS PRACTICE

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. The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b. $\boxed{a = 62.2}$ $\boxed{b = 18.8}$

(ii) Write down the value of 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) answer part (c).

[1]

c) Sub $x = 8$ into y

$y = 62.2(8) + 18.8$

$y = 516.4285...$

$\boxed{y = 516 \text{ calories } (3sf)}$

N.B calculator values for a and b used.

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. The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b.

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

$\boxed{r = 0.991 \ (3sf)}$

[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) answer part (c).

[1]

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

The relationship between A and V can be modelled by the regression line of V on A with equation $V = aA + b$.

(a) (i) Find the value of a and the value of b.

(ii) Write down the value of 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) $a = 0.6742...$ $b = 38.3809...$

$\boxed{a = 0.674 \ (3sf)}$ $\boxed{b = 38.4 \ (3sf)}$

ii) $r = 0.9437...$

$\boxed{r = 0.944 \ (3sf)}$

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

The relationship between A and V can be modelled by the regression line of V on A with equation $V = aA + b$.

(a) (i) Find the value of a and the value of b.

$\boxed{a = 0.674}$

(ii) Write down the value of 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

The relationship between A and V can be modelled by the regression line of V on A with equation $V = aA + b$.

(a) (i) Find the value of a and the value of b. $\boxed{a = 0.674}$ $\boxed{b = 38.4}$

(ii) Write down the value of 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]

c) Sub $A = 95$ into V.

$V = 0.674(95) + 38.4$

$V = 102.4380...$

$\boxed{V = 102\% \quad (3sf)}$

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

The relationship between A and V can be modelled by the regression line of V on A with equation $V = aA + b$.

(a) (i) Find the value of a and the value of b.

(ii) Write down the value of 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.

$\boxed{V = 102\% \quad (3sf)}$

[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

Let L_1 be the regression line of P on D. The equation of the line L_1 can be written in the form $P = aD + b$.

Let L_2 be the regression line of D on P. The equation of the line L_2 can be written in the form $D = cP + d$.

(a) (i) Find the value of a and the value of b.

(ii) Find the value of c and the value of d.

[3]

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

[2]

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

[2]

The lines L_1 and L_2 both pass through the same point with coordinates (p, q).

(d) Find the value of p and the value of q.

[2]

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

i) x list: distance

y list: price

$a = 0.06289\ldots$ $\quad$ $b = 29.0623\ldots$

$\boxed{a = 0.0629 \text{ (3sf)}}$ $\quad$ $\boxed{b = 29.1 \text{ (3sf)}}$

ii) x list: price

y list: distance

$c = 14.760\ldots$ $\quad$ $d = -370.397\ldots$

$\boxed{c = 14.8 \text{ (3sf)}}$ $\quad$ $\boxed{d = -370 \text{ (3sf)}}$

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

Let L_1 be the regression line of P on D. The equation of the line L_1 can be written in the form $P = aD + b$.

Let L_2 be the regression line of D on P. The equation of the line L_2 can be written in the form $D = cP + d$.

(a) (i) Find the value of a and the value of b.

(ii) Find the value of c and the value of d.

[3]

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

[2]

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

[2]

The lines L_1 and L_2 both pass through the same point with coordinates (p, q).

(d) Find the value of p and the value of q.

[2]

b) Use the regressions from part (a).

$r = 0.9634\ldots$

$\boxed{r = 0.963 \text{ (3sf)}}$

N.B r is the same for both regressions.

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

Let L_1 be the regression line of P on D. The equation of the line L_1 can be written in the form $P = aD + b$.

Let L_2 be the regression line of D on P. The equation of the line L_2 can be written in the form $D = cP + d$.

(a) (i) Find the value of a and the value of b. $\boxed{a = 0.0629}$ $\boxed{b = 29.1}$

(ii) Find the value of c and the value of d.

[3]

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

[2]

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

[2]

The lines L_1 and L_2 both pass through the same point with coordinates (p, q).

(d) Find the value of p and the value of q.

[2]

c) Sub $D = 2635$ into P.

$P = 0.0629(2635) + 29.1$

$P = 194.7836...$

$\boxed{P = 195 \text{ US dollars (3sf)}}$

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

Let L_1 be the regression line of P on D. The equation of the line L_1 can be written in the form $P = aD + b$.

Let L_2 be the regression line of D on P. The equation of the line L_2 can be written in the form $D = cP + d$.

(a) (i) Find the value of a and the value of b. $\boxed{a = 0.0629}$ $\boxed{b = 29.1}$

(ii) Find the value of c and the value of d. $\boxed{c = 14.8}$ $\boxed{d = -370}$

[3]

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

[2]

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

[2]

The lines L_1 and L_2 both pass through the same point with coordinates (p, q).

(d) Find the value of p and the value of q.

[2]

d) Solve simultaneous equations on your GDC.

$L_1: P = 0.0629D + 29.1$

$L_2: D = 14.8P - 370$

Use calculator values for a, b, c and d.

$D = p = 816.875...$ $P = q = 80.4375...$

$\boxed{p = 817 \text{ (3sf)}}$ $\boxed{q = 80.4 \text{ (3sf)}}$