Exam Papers Practice

4.2 Correlation & Regression

Mark Schemes

Course	DP IB Maths
Section	4. Statistics & Probability
Topic	4.2 Correlation & Regression
Difficulty	Medium

To be used by all students preparing for DP IB Maths AA SL
Students of other boards may also find this useful

Question 1

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

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

Question 2

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

x list: T

y list: C

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

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

ii) $r = -0.9425...$

$r = -0.942$ (3sf)

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.

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

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

x list: age

y list: height

i) $a = 5.8757...$ $b = 78.7259...$

$a = 5.88$ (3sf) **$b = 78.7$ (3sf)**

ii) $r = 0.9843...$

$r = 0.984$ (3sf)

b) Sub $x = 9$ into y.

$y = 5.88(9) + 78.7$

$y = 131.6079...$

$y = 132$ cm (3sf)

N.B calculator values for a and b used.

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

Question 4

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

x list: distance

y list: calories

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

$a = 62.2$ (3sf) $b = 18.8$ (3sf)

ii) $r = 0.9907...$

$r = 0.991$ (3sf)

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

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).

Question 5

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

x list: age

y list: percentage of willing people

i) $a = 0.6742...$ $b = 38.3809...$

$a = 0.674$ (3sf) $b = 38.4$ (3sf)

ii) $r = 0.9437...$

$r = 0.944$ (3sf)

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

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.

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

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

i) x list: distance

y list: price

$a = 0.06289...$ $b = 29.0623...$

$a = 0.0629$ (3sf) $b = 29.1$ (3sf)

ii) x list: price

y list: distance

$c = 14.760...$ $d = -370.397...$

$c = 14.8$ (3sf) $d = -370$ (3sf)

b) Use the regressions from part (a).

$r = 0.9634...$

$r = 0.963$ (3sf)

N.B r is the same for both regressions.

c) 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.

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.

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

$p = 817$ (3sf) **$q = 80.4$ (3sf)**