



eClassroom

A-LEVEL MATHEMATICS • STATISTICS 1

Correlation and Regression

Worked Solutions

Pearson Edexcel International A Level (WST01) & A Level (9MA0) • Calculator and statistical tables allowed: show every step

Full worked solutions for the free Correlation and Regression worksheet, A-Level and International A-Level Maths



Tap a section to jump straight to it

Full worked solutions for every question on the Statistics 1 Correlation and Regression worksheet, in the same order and with the same question numbers.

How to mark your work. Tick each step you got right, not just the final answer. At A-level most marks are for method, so a wrong final answer can still earn most of the marks. If you lost marks, find the *first* line where your working is different from ours. That is the thing to fix.

In Section E: **M** = method mark, **dM** = method mark that depends on the previous M mark, **A** = accuracy mark (needs the M mark before it), **B** = independent mark.

Question sheet

Exam questions by topic

Statistics 1 exam questions

More worksheets



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

Page 1 of 7

A**Your turn answers**

Compare each line of your working with the solution.

CORE

A1 $S_{xx} = 90 - 80 = 10$, $S_{yy} = 265 - 245 = 20$, $S_{xy} = 150 - 140 = 10$: $\frac{10}{\sqrt{200}} \approx \boxed{0.707}$

A2 $b = \frac{-40}{25} = -1.6$, $a = 30 + 6.4$: $\boxed{y = 36.4 - 1.6x}$

A3 **weak positive correlation**: larger cities tend to have slightly more rainfall, but the link is weak.

STRETCH

A4 $x = 5$: $\boxed{28.4}$, interpolation, reliable; $x = 25$: $\boxed{-3.6}$, extrapolation, and a negative value is impossible here

A5 $y - 100 = 5 + 1.2 \left(\frac{x - 20}{2} \right) = 5 + 0.6x - 12$: $\boxed{y = 93 + 0.6x}$

A6 $S_{ww} = \frac{1}{4}S_{dd}$ and $S_{tw} = \frac{1}{2}S_{td}$: $\boxed{S_{dd} = 144, S_{td} = -24}$

A7 $0.8 \times (\text{increase}) = 12$: $\boxed{15}$

B**Fluency solutions**

One mark for each correct part. The notes in italics explain why each sequence was chosen.

CORE

B1 $\sum x = 27$, $\sum y = 40$, $\sum x^2 = 175$, $\sum y^2 = 390$, $\sum xy = 261$

(a) $175 - \frac{27^2}{5}$: $\boxed{29.2}$ (b) $390 - 320$: $\boxed{70}$ (c) $261 - 216$: $\boxed{45}$ (d) $\frac{45}{\sqrt{29.2 \times 70}} \approx \boxed{0.995}$

The three S values come from the same five sums.



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

B2 (a) $\frac{45}{29.2} \approx \boxed{1.54}$ (b) $8 - 1.541 \times 5.4 \approx \boxed{-0.322}$ (c) $\boxed{8.92}$

(d) y increases by about 1.54 for each increase of 1 in x

Keep b unrounded when finding a

B3 (a) **strong positive** (b) **weak negative** (c) **impossible**: $-1 \leq r \leq 1$

(d) **very strong negative**

Always check $|r| \leq 1$

B4 (a) $S_{xx} = 40, S_{yy} = 350, S_{xy} = 80$ (b) $\frac{80}{\sqrt{14000}} \approx \boxed{0.676}$ (c) $\boxed{2}$ (d) $a = 25 - 12: \boxed{y = 13 + 2x}$

$\bar{x} = 6$ and $\bar{y} = 25$ come from the sums.

CORE

B5 (a) $\boxed{27}$ (b) $\boxed{63}$ (c) $\boxed{(a)}$: $x = 7$ is inside the data; 25 is far outside

Interpolation is reliable; extrapolation is not.

B6 (a) **hours of revision, h** (b) **the line of m on h**

(c) **it minimises errors in m , not in h** (d) **correlation does not show cause**: other factors could explain both

The explanatory variable goes on the right of the line.

STRETCH

B7 (a) $\boxed{0.9}$ (b) $\frac{y-50}{4} = 1.2 + 0.8(x - 100): \boxed{y = -265.2 + 3.2x}$ (c) $\boxed{86.8}$

(d) **linear coding with positive scales does not change r**

Decode the line by substituting both codings.

B8 (a) $\frac{900}{25}: \boxed{36}$ (b) $\frac{120}{5}: \boxed{24}$ (c) $\frac{24}{\sqrt{20 \times 36}} \approx \boxed{0.894}$ (d) $\frac{24}{20}: \boxed{1.2}$

S_{ww} scales by 5^2 ; S_{tw} by 5

B9 (a) $0.6 \times 10: \boxed{6g}$ (b) $\frac{15}{0.6}: \boxed{25cm}$ (c) $-8 + 0.6 \times 120: \boxed{64g}$

(d) $-8 + 6 = -2g: \boxed{\text{impossible, so the model fails for very small plants}}$

Convert to the units of the data before substituting.



- B10 (a) $\bar{x} = 6, \bar{y} = 25$ (b) unchanged: 40, 350, 80 (c) 0.676 (d) $y = 13 + 2x$

The new point adds nothing to any $(x - \bar{x})$ or $(y - \bar{y})$ term.

C

Reasoning solutions

Your wording can be different. Check that your reason makes the same point.

CORE

- C1 It needs context: on warmer days, more cold drinks tend to be sold

- C2 $y = 5.5$ is in hundreds of dollars: \$550

- C3 $|r|$ cannot exceed 1; he divided by S_{xx} , which gives the gradient. $\frac{54}{\sqrt{50 \times 72}} = \frac{54}{60} = 0.9$

- C4 $S_{ww} = 4^2 S_{dd}$, so divide by 16: $S_{dd} = 20$

STRETCH

- C5 $25 - 1.5 \times 20 = -5$ minutes: a negative waiting time is impossible, so the model fails there.

- C6 A change in g uses the gradient only: $0.4 \times (\text{increase}) = 6$: 15 cm

- C7 The line of y on x minimises errors in y for given x ; it is not designed to estimate x from y

- C8 (a) $a = \bar{y} - b\bar{x}$, so at $x = \bar{x}$: $y = \bar{y} - b\bar{x} + b\bar{x} = \bar{y}$: the line passes through $(\bar{x}, \bar{y})$

- (b) $b = \frac{S_{xy}}{S_{xx}}$ and $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$, with $S_{xx} > 0$ and $S_{yy} > 0$: both have the sign of S_{xy}



D**Problem solving solutions**

There is often more than one correct method. Any complete, correct method scores full marks.

CORE

D1 (a) $39\,400 - \frac{1240^2}{40} = 39\,400 - 38\,440$: **960**

(b) $\frac{108}{\sqrt{960 \times 96}} \approx$ **0.356**: weak positive, so taller plants tend to have slightly longer leaves

D2 (a) $950 - \frac{52 \times 140}{8}$: **$S_{xy} = 40$** ; $2530 - \frac{140^2}{8}$: **$S_{yy} = 80$**

(b) $\frac{40}{\sqrt{42 \times 80}} \approx$ **0.690**: moderate positive, so players who train more tend to score more goals

D3 **r is unchanged**: multiplying by a positive constant does not change the correlation.

The gradient is divided by 1.6: each kilometre is a smaller step than a mile.

The intercept is unchanged: $x = 0$ is still distance 0

D4 **$r = 0.94$** , unchanged by coding.

$h = 2y + 20$, $t = x + 10$: $2y + 20 = 12.4 + 0.86(x + 10)$, so $2y = 1 + 0.86x$: **$y = 0.5 + 0.43x$**

STRETCH

D5 (a) $S_{td} = \frac{-224}{8} = -28$; $S_{dd} = \frac{1600}{64} = 25$

$\frac{-28}{\sqrt{40 \times 25}} \approx$ **-0.885**

(b) $8d = 50 - 5.6t$: **$d = 6.25 - 0.7t$**

D6 (a) $1.8 \text{ m} = 180 \text{ cm}$: $-30 + 0.65 \times 180$: **87 cm**

(b) $0.65x = 13$: **$x = 20$**

D7 $x = 0.4(2x + 3) + 0.6 = 0.8x + 1.8$, so $0.2x = 1.8$: **$\bar{x} = 9$, $\bar{y} = 21$**

10×9 and 10×21 : **$\sum x = 90$, $\sum y = 210$**



- D8 (a) $\bar{f} = \frac{744}{12} = 62$; $\sqrt{\frac{46548}{12} - 62^2} = \sqrt{35} \approx$ **mean 62 °F, sd 5.92 °F**
- (b) $\frac{5}{9}(62 - 32)$ and $\frac{5}{9} \times 5.92$: **mean 16.7 °C, sd 3.29 °C**
- (c) **0.83**: a linear coding with a positive multiplier does not change r

E

Exam practice solutions

Mark-scheme notes are shown on the right of each line.

CORE

Exam questions for everyone

- E1 (a) **to simplify a real-world situation** B1
- (b) $\frac{315}{\sqrt{52.5 \times 2016}}$ M1
- 0.968** A1
- (c) **strong positive: sunnier days generate more electricity** B1
- (d) $b = \frac{315}{52.5} = 6$; $\bar{s} = 7$, $\bar{y} = 44$; $a = 44 - 6 \times 7 = 2$ M1 M1
- $y = 2 + 6s$** A1*
- (e) $10e = 2 + 6s$: **$e = 0.2 + 0.6s$** B1
- (f) $0.2 + 0.6 \times 9$: **5.6 kWh**; **each extra hour of sunshine adds about 0.6 kWh** B1 B1
- E2 (a) $S_{dh} = 8320 - \frac{450 \times 180}{10}$ M1
- 220**; $S_{hh} = 3340 - \frac{180^2}{10}$: **100** A1 A1
- (b) $\frac{220}{\sqrt{1250 \times 100}}$ M1
- 0.622** A1
- (c) $b = \frac{220}{1250} = 0.176$; $a = 18 - 0.176 \times 45 = 10.08$ M1 M1
- $h = 10.08 + 0.176d$** A1*
- (d) **each extra 1 cm of diameter goes with about 0.176 m more height** B1
- (e) $0.5 \text{ m} = 50 \text{ cm}$: $10.08 + 0.176 \times 50$ M1
- 18.9 m** A1



STRETCH

- E3** (a) Ordered marks: $Q_1 = 44$ (3rd), $Q_3 = 62$ (8th), IQR 18 M1
 $62 + 1.5 \times 18 = 89$ M1
 $95 > 89$: **an outlier** A1
- (b) $S_{xx} = 615 - \frac{69^2}{9} = 86$, $S_{xy} = 3761 - \frac{69 \times 451}{9} = 303.33$ M1
 $b = \frac{303.33}{86} = 3.527$, $a = \frac{451}{9} - 3.527 \times \frac{69}{9} = 23.07$ M1
 $y = 23.1 + 3.53x$ A1*
- (c) **each extra hour of revision goes with about 3.5 more marks** B1
- (d) $x = 7.5$: **49.5**, reliable (interpolation) B1
 $x = 30$: **129**, unreliable: extrapolation, and over 100 is impossible M1 A1
- (e) **less**: the outlier lies far from the linear pattern B1
- (f) 5×3.53 M1
about 17.6 marks A1

