Unit 10 · Correlation and Regression

ANSWER KEY · FOR TEACHERS

Full marks 27

Give the marks for the points in the marking scheme; an answer in other words that says the same is right.

1 Key words

K · 4

1. bivariate
2. scatter diagram
3. negative
4. straight line
5. rank
6. least squares
7. coefficient
8. means

Each carries 0.5 mark.

2 True or false

U · 3

i. ✗ **False**: $r = 0$ means no **straight-line** relation: points on $y = x^2$ can give $r = 0$, though y depends on x exactly.

ii. ✓ **True**

iii. ✗ **False**: Use the line of x on y; solving the y-on-x line for x gives a different answer unless $r = \pm 1$.

iv. ✓ **True**

v. ✗ **False**: Correlation is not cause: a third thing can drive both (ice cream sales and drownings both rise in a hot month).

vi. ✓ **True**

Each carries 0.5 mark.

3 Multiple choice

i. K · 1

a. $r\frac{\sigma_y}{\sigma_x}$

$b_{yx} = r\frac{\sigma_y}{\sigma_x}$: how much y changes for each unit of x. $r\frac{\sigma_x}{\sigma_y}$ is b_{xy}.

ii. U · 1

c. -0.98 and -0.5

Both must have the sign of r, and their product must be $r^2 = 0.49$: $(-0.98)(-0.5) = 0.49$. -1.4 and -0.5 give 0.7, which is $|r|$, not r^2.

iii. U · 1

c. 0.8

$n\Sigma xy - \Sigma x\Sigma y = 680 - 600 = 80$; $n\Sigma x^2 - (\Sigma x)^2 = 275 - 225 = 50$; $n\Sigma y^2 - (\Sigma y)^2 = 1800 - 1600 = 200$.

$r = \frac{80}{\sqrt{50 \times 200}} = \frac{80}{100} = 0.8$. 1.6 is $b_{yx} = \frac{80}{50}$.

iv. A · 1

b. 36

$b_{yx} = 0.6 \times \frac{4}{2} = 1.2$, and $y = 30 + 1.2(15 - 10) = 36$. 31.5 uses $b_{xy} = 0.6 \times \frac{2}{4} = 0.3$; 42 uses $\frac{\sigma_y}{\sigma_x} = 2$ without r.

Science Playground, Katunje, Bhaktapur
Book a class visit: https://science-playground.com/booking

v. **c.** becomes half HA · 1

Adding 5 to y is a change of origin and changes nothing in b_{yx}. Doubling x doubles σ_x and leaves r as it was, so $b_{yx} = r\frac{\sigma_y}{\sigma_x}$ halves: y now changes by the same amount over twice as many units of x. Only r has no units to change.

4 Short answers

a. MARKING SCHEME U · 2

✓ $b_{yx}\ b_{xy} = r^2$.

✓ $r^2 \leq 1$, so both can't be more than 1 in size.

b. MARKING SCHEME U · 2

✓ No: a correlation does not show that one causes the other.

✓ A third thing, the rain (the monsoon months), drives both.

c. MARKING SCHEME A · 3

✓ The differences d and $\Sigma d^2 = 12$.

✓ $\rho = 1 - \frac{72}{210} = 0.66$.

✓ A moderate positive agreement between the judges.

5 Long answer

ANSWER A · 5

a. $\bar{x} = \frac{25}{5} = 5, \bar{y} = \frac{150}{5} = 30$. With $X = x - 5$, $Y = y - 30$: $X = -3, -1, 0, 1, 3$; $Y = -10, -4, 1, 4, 9$.

$\Sigma XY = 30 + 4 + 0 + 4 + 27 = 65$, $\Sigma X^2 = 9 + 1 + 0 + 1 + 9 = 20$, $\Sigma Y^2 = 100 + 16 + 1 + 16 + 81 = 214$.

$r = \frac{65}{\sqrt{20 \times 214}} = \frac{65}{65.42} = 0.99$: a very high positive correlation.

b. $b_{yx} = \frac{65}{20} = 3.25$: $y - 30 = 3.25(x - 5)$, that is $y = \mathbf{13.75 + 3.25}x$. At $x = 7$: $y = 30 + 3.25 \times 2 = \mathbf{36.5}$ **words a minute**.

What it means: each extra hour of practice a week goes with about 3 more words a minute, among these five students.

MARKING SCHEME

✓ The means 5 and 30 and the deviations.

✓ $\Sigma XY = 65$, $\Sigma X^2 = 20$, $\Sigma Y^2 = 214$.

✓ $r = 0.99$.

✓ $b_{yx} = 3.25$ and the line $y - 30 = 3.25(x - 5)$.

✓ 36.5 words a minute.

6 Think

ANSWER HA · 3

$x = 88.49$: $y = 25.45 - 0.487 \times 88.49 = 25.45 - 43.09 \approx \mathbf{-17.6\ °C}$.

The line was found from stations between about 70 m and 2,400 m; it says how temperature and height go together only in that range. Nothing in the data says the same straight line goes on to 8,849 m, more than three times higher than any station, so the number is a guess outside the data, not an estimate.

MARKING SCHEME

✓ $y = 25.45 - 0.487 \times 88.49 \approx -17.6$ °C.

✓ The data stop near 2,400 m: estimate only inside the range of the data.

✓ Nothing shows the relation stays a straight line far beyond it.

Science Playground, Katunje, Bhaktapur
Book a class visit: https://science-playground.com/booking

7 Activity: Arm span and height

(not marked)

WHAT STUDENTS SHOULD FIND

The points rise to the right close to a line: a high positive correlation, usually r above 0.8 (with friends of nearly one height, the spread is small and r can come out lower). The arm span is usually within a few centimeters of the height, so the line y on x has a slope near 1 and passes close to $y = x$.

Marks by level	Knowledge	Understanding	Application	Higher ability	Total
Marks	5	9	9	4	**27**

Science Playground, Katunje, Bhaktapur
Book a class visit: https://science-playground.com/booking