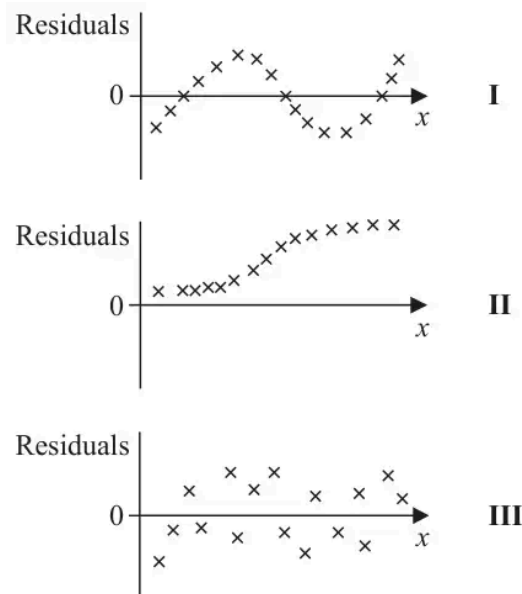


QQQ – Further Statistics 2 – Chapter 1 – Linear Regression (Version 1)

Total Marks: 25

(25 = Platinum, 23 = Gold, 20 = Silver, 17 = Bronze)

1. Below are 3 sketches from some students of the residuals from their linear regressions of y on x .



For each sketch you should state, giving your reason,

- (i) whether or not the sketch is feasible

and if it is feasible

- (ii) whether or not the sketch suggests a linear or a non-linear relationship between y and x .

(6)

(Total 6 marks)

2. Two students are experimenting with some water in a plastic bottle. The bottle is filled with water and a hole is put in the bottom of the bottle. The students record the time, t seconds, it takes for the water level to fall to each of 10 given values of the height, h cm, above the hole.

Student A models the data with an equation of the form $t = a + b\sqrt{h}$

The data is coded using $v = t - 40$ and $w = \sqrt{h}$ and the following information is obtained.

$$\sum v = 626 \quad \sum v^2 = 64\,678 \quad \sum w = 22.47 \quad S_{ww} = 4.52 \quad S_{vw} = -338.83$$

(a) Find the equation of the regression line of t on $\sqrt{h}$ in the form $t = a + b\sqrt{h}$ **(4)**

The time it takes the water level to fall to a height of 9 cm above the hole is 47 seconds.

(b) Calculate the residual for this data point.
Give your answer to 2 decimal places. **(2)**

Given that the residual sum of squares (RSS) for the model of t on $\sqrt{h}$ is the same as the RSS for the model of v on w ,

(c) calculate the RSS for these 10 data points. **(2)**

Student B models the data with an equation of the form $t = c + dh$

The regression line of t on h is calculated and the residual sum of squares (RSS) is found to be 980 to 3 significant figures.

(d) With reference to part (c) state, giving a reason, whether Student B 's model or Student A 's model is the more suitable for these data. **(1)**

(Total 9 marks)

3. A medical researcher is exploring the relationship between the weight, x kg, and the head circumference, y cm, of newborn babies. A random sample of 15 newborn babies is taken and the data are summarised by the following statistics

$$\sum x = 50.46 \quad \sum x^2 = 171.828 \quad \sum y = 518.9 \quad \sum y^2 = 18\,004.47 \quad S_{xy}$$

(a) Show that, to 3 decimal places, $S_{xx} = 2.081$ and $S_{yy} = 53.989$ **(3)**

(b) Find the equation of the regression line of y on x , giving your answer in the form $y = a + bx$ **(3)**

(c) Find the residual sum of squares (RSS) for these data. **(1)**

One of these 15 babies had a birth weight of 3.26 kg and a head circumference of 36.8 cm

(d) Find the residual for this baby. **(2)**

The researcher claims that this baby could be an outlier.

(e) Using your answers to part (c) and part (d), give a reason that might support this claim. **(1)**

(Total 10 marks)

Solutions

1.

Scheme	Marks	AO
I (Is feasible as a residual plot but) probably a non-linear relationship	B1	2.2b
Since the residuals are not randomly scattered about zero	B1	2.4
II Impossible as a residual plot	B1	2.2a
Since the residuals do not sum to zero	B1	2.4
III (Is feasible as a residual plot) and probably a linear relationship	B1	2.2b
Since the points are randomly scattered about zero	B1	2.4
	(6)	
	(6 marks)	

Notes

I: 1st B1 for stating possibly non-linear (allow a suitable sketch)

2nd B1 for a suitable comment (e.g. follow a systematic pattern)

II: 1st B1 for stating not feasible as a residual plot

2nd B1 for a correct reason

III: 1st B1 for stating probably a linear relationship

2nd B1 for a suitable supporting reason

2.

Mark scheme (a)

Scheme	Marks	AO
$b = \frac{-338.83}{4.52} [= -74.96 \dots]$	M1	3.3
$a = \frac{626}{10} - "b" \frac{22.47}{10} [= 231.04 \dots]$	M1	1.1b

$t - 40 = "231.04 \dots" + (" - 74.96 \dots") \sqrt{h}$	dM1	3.1a
$t = 271.04 \dots - 74.96 \dots \sqrt{h}$	A1	1.1b
	(4)	

Notes

M1: For use of a correct model ie a correct expression for b

M1: For use of a correct model ie a correct expression (ft) for a

dM1: dep on both previous method marks for proceeding from an equation of the form $v = "a" + "b"w$ to a correct un-simplified model in terms of h and t ft their a and b

A1: For a correct model $t = 271.04 \dots - 74.96 \dots \sqrt{h}$ with awrt 271 and awrt 75 (corrected from the printed mark scheme: the note prints 74.95...; $b = -74.96 \dots$)

Mark scheme (b)

Scheme	Marks	AO
Residual = $47 - ("271.04 \dots" - "74.96 \dots" \times \sqrt{9})$	M1	3.4
= 0.8466 ...	A1	1.1b
	(2)	

Notes

M1: For a correct method to find the residual. States $h = 9$ and a correct expression to find the residual for their model of the form $t = a + b\sqrt{h}$ or 9 substituted into a correct expression. Must be subtracting the correct way round. Alternatively, may code the data to v and w and attempt

$$47 - ("231.04 \dots" - "74.96 \dots" \times 3)$$

A1: awrt 0.85 Allow answers in range awrt 0.84 to awrt 0.86 if working shown following a correct model in (a). If no working is shown then look for awrt 0.847

Mark scheme (c)

Scheme	Marks	AO
$\text{RSS} = \left[64\,678 - \frac{626^2}{10} \right] - \frac{(-338.83)^2}{4.52} \left(= 25\,490.4 - \frac{(-338.83)^2}{4.52} \right)$	M1	1.1b
= 90.89 ... [s ²]	A1	1.1b

	(2)	
--	-----	--

Notes

M1: For a correct expression for RSS. 25490.4 may be seen as $\frac{127\,452}{5}$

May also use $RSS = S_{vv}(1 - r^2) = 25\,490.4 \left(1 - \frac{(-338.83)^2}{4.52 \times 25\,490.4} \right)$

A1: awrt 90.9

Mark scheme (d)

Scheme	Marks	A0
Student <i>A</i> 's model as the sum of squares of the residuals is lower	B1	2.4
	(1)	
	(9 marks)	

Notes

B1: Explaining a reason **for their conclusion that A is a more suitable model** provided their positive RSS found in (c) is less than 980. e.g. RSS is smaller so model *A*

Condone references to the model being more accurate oe. Must be a comparison with *B* or implied so do not accept statements such as “*A* because it has a small RSS”

3.

Mark scheme (a)

Scheme	Marks	A0
$S_{xx} = 171.828 - \frac{50.46^2}{15}$ or $S_{yy} = 18\,004.47 - \frac{518.9^2}{15}$	M1	1.1b
$= 2.08056 = \underline{2.081}^* (3dp)$; $= 53.98933 \dots = \underline{53.989}^* (3dp)$	A1*; A1*	1.1b(x2)
	(3)	

Notes

M1: for either correct expression

1st A1*: cso for 2.081 or better following a correct expression

2nd A1*: cso for 53.989 or better following a correct expression

Mark scheme (b)

Scheme	Marks	AO
$b = \left[\frac{S_{xy}}{S_{xx}} = \right] \frac{7.8284}{2.081} = (3.76 \dots)$	M1	1.1b
$a = \frac{518.9}{15} - b \frac{50.46}{15} = 34.593 \dots - "3.76 \dots" \times 3.364 (= 21.9 \dots)$	M1	1.1b
Equation of line is: $y = 21.9 + 3.76x$	A1	1.1b
	(3)	

Notes

1st M1: for a correct numerical expression for gradient (or ft their $S_{xx} \neq 2.081, S_{xy} \neq 7.8284$)

2nd M1: for a correct expression for intercept (ft their b). May be implied by awrt 21.9

A1: for equation with $a =$ awrt 21.9 and $b =$ awrt 3.76

Note: A correct equation scores M1M1A1

Mark scheme (c)

Scheme	Marks	AO
$\left[\text{RSS} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = \right] 24.5397 \dots = \text{awrt } \underline{24.5}$	B1	1.1b
	(1)	

Notes

B1: for awrt 24.5

Mark scheme (d)

Scheme	Marks	AO
$\hat{y} = "21.9" + "3.76" \times 3.26 (= 34.1576)$	M1	3.4
[Residual = $y - \hat{y} = 36.8 - "34.15 \dots" = 2.6424$] = awrt <u>2.6</u>	A1	1.1b
	(2)	

Notes

M1: for an expression for $\hat{y}$ or the residual (ft their regression line equation). Implied by awrt 2.6

A1: for awrt 2.6 (exact figures give 2.5979...)

Mark scheme (e)

Scheme	Marks	AO
"2.64 . . ." ≈ 7 which is a high proportion of RSS (so this value may be an outlier)	B1	2.4
	(1)	
	(10 marks)	

Notes

B1: Dependent on $0.25 < \frac{2.6^2}{24.5} < 1$ for a suitable explanation based on their (d)² compared with their (c) or other suitable calculations which enable a comparison to be made.
e.g. this residual is 27.5% of the total which is a large proportion".
References to size e.g. "large" requires quoting the numerical value as well.