Correlation - exam questions

The head and body length, x millimetres, and tail length, y millimetres, of each of a sample of 20 adult dormice were measured. The following statistics are derived from the results.

$$S_{xx} = 1280.55 \qquad S_{yy} = 281.8 \qquad S_{xy} = 416.3$$

(a) Calculate the value of the product moment correlation coefficient between x and y.

(2 marks)

(b) Interpret your value in the context of this question. *(2 marks)*

(c) Write down the value of the product moment correlation coefficient if the measurements had been recorded in centimetres. *(1 mark)*

(d) Give a reason why it is not generally advisable to calculate the value of the product moment correlation coefficient without first viewing a scatter diagram of the data. Illustrate your answer with a sketch. *(2 marks)*
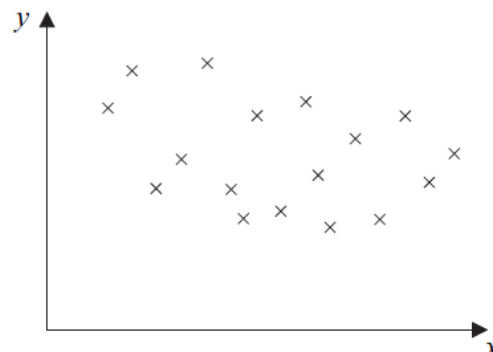
1 (a) Estimate, **without undertaking any calculations**, the value of the product moment correlation coefficient between the variables x and y for each of the two scatter diagrams.

(i)   **(ii)**

(2 marks)

(b) The table gives the circumference, x centimetres, and the weight, y grams, of each of 12 new cricket balls.

x	22.5	22.7	22.6	22.4	22.5	22.8	22.6	22.7	22.8	22.4	22.9	22.6
y	160.3	159.4	157.8	158.0	157.3	159.8	158.3	159.6	161.3	156.4	162.5	161.2

(i) Calculate the value of the product moment correlation coefficient between x and y.

(3 marks)

(ii) Assuming that the 12 balls may be considered to be a random sample, interpret your value in context. *(2 marks)*

Hermione, who is studying reptiles, measures the length, x cm, and the weight, y grams, of a sample of 11 adult snakes of the same type. Her results are shown in the table.

Snake	A	B	C	D	E	F	G	H	I	J	K
x	46	39	54	79	47	58	73	35	43	51	36
y	55	48	58	88	61	55	82	51	50	66	57

(a) Calculate the value of the product moment correlation coefficient, r, between x and y. *(3 marks)*

(b) Interpret your value in context. *(2 marks)*

(c) Complete the scatter diagram, opposite, for these data. *(2 marks)*

(d) Subsequently it is found that, of the 11 adult snakes, 9 are male and 2 are female.

   (i) Given that female adult snakes are generally larger than male adult snakes, identify the 2 snakes which are most likely to be female. *(1 mark)*

   (ii) Hence, **without further calculation**, estimate the value of r for the 9 male snakes and revise, as necessary, your interpretation in part **(b)**. *(2 marks)*

A greengrocer sells bunches of 9 carrots at his Saturday market stall. Tom and Geri are two Statistics students who work on the stall. Each selects a bunch of carrots at random.

(a) At home, Tom measures the length, x centimetres, and the maximum diameter, y centimetres, of each carrot in his selected bunch with the following results.

x	16.2	13.1	10.4	12.1	14.6	9.7	11.8	13.6	17.3
y	4.2	3.9	4.7	3.3	3.7	2.4	3.1	3.5	2.7

   (i) Calculate the value of the product moment correlation coefficient. *(3 marks)*

   (ii) Interpret your value in context. *(2 marks)*

(b) At her home, Geri measures the length, in centimetres, and the **weight**, in grams, of each carrot in her selected bunch and then obtains a value of -0.986 for the product moment correlation coefficient.

   Comment, with a reason, on the likely validity of Geri's value. *(2 marks)*

The table shows the length, in centimetres, and maximum diameter, in centimetres, of each of 10 honeydew melons selected at random from those on display at a market stall.

Length	24	25	19	28	27	21	35	23	32	26
Maximum diameter	18	14	16	11	13	14	12	16	15	14

(a)   Calculate the value of the product moment correlation coefficient.   *(3 marks)*

(b)   Interpret your value in the context of this question.   *(2 marks)*

The table shows, for each of a random sample of 8 paperback fiction books, the number of pages, x, and the recommended retail price, $£y$, to the nearest 10p.

x	223	276	374	433	564	612	704	766
y	6.50	4.00	5.50	8.00	4.50	5.00	8.00	5.50

(a)   (i)   Calculate the value of the product moment correlation coefficient between x and y.
          (3 marks)

(ii)   Interpret your value in the context of this question.   *(2 marks)*

(iii)   Suggest one other variable, in addition to the number of pages, which may affect the recommended retail price of a paperback fiction book.   *(1 mark)*

(b)   The same 8 books were later included in a book sale. The value of the product moment correlation coefficient between the number of pages and the sale price was 0.959, correct to three decimal places.

What can be concluded from this value?   *(2 marks)*

7 (a)   Three airport management trainees, Ryan, Sunil and Tim, were each instructed to select a random sample of 12 suitcases from those waiting to be loaded onto aircraft.

Each trainee also had to measure the volume, x, and the weight, y, of each of the 12 suitcases in his sample, and then calculate the value of the product moment correlation coefficient, r, between x and y.

- Ryan obtained a value of -0.843.
- Sunil obtained a value of $+0.007$.

Explain why neither of these two values is likely to be correct.   *(2 marks)*

(b)   Peggy, a supervisor with many years' experience, measured the volume, x cubic feet, and the weight, y pounds, of each suitcase in a random sample of 6 suitcases, and then obtained a value of 0.612 for r.

- Ryan and Sunil each claimed that Peggy's value was different from their values because she had measured the volumes in cubic feet and the weights in pounds, whereas they had measured the volumes in cubic metres and the weights in kilograms.
- Tim claimed that Peggy's value was almost exactly half his calculated value because she had used a sample of size 6 whereas he had used one of size 12.

Explain why neither of these two claims is valid.   *(2 marks)*

(c)   Quentin, a manager, recorded the volumes, v, and the weights, w, of a random sample of 8 suitcases as follows.

v	28.1	19.7	46.4	23.6	31.1	17.5	35.8	13.8
w	14.9	12.1	21.1	18.0	19.8	19.2	16.2	14.7

(i)   Calculate the value of r between v and w.   *(3 marks)*

(ii)   Interpret your value in the context of this question.   *(2 marks)*

Question 1: Jan 2008

2(a)	$r = \dfrac{416.3}{\sqrt{1280.55 \times 281.8}} =$	M1		Allow no $\sqrt{}$	
	0.69 to 0.7(0)	A1	2	AWFW $\qquad\qquad$ (0.693) (0.00115)	
(b)	(Quite or fairly) **weak** / **some** / **moderate** (quite or fairly) **strong positive correlation** (relationship / association)	A1		OE; must qualify strength and indicate positive A0 for poor / reasonable / average / medium / good A0 for very weak / very strong etc	
	between				
	head & body length and **tail length**	B1	2	Context; accept 'body and tail' or even 'head and tail'	
	Ignore subsequent alternative comments only if A1 B1 already scored				
	OR				
	Some evidence that mice with large head & body lengths also have long tails	(A1) (B1)		OE; must qualify strength and indicate positive in context	
(c)	0.69 to 0.7(0) **OR** Answer to (a)	B1✓	1	$0 < r < 1$	
(d)	Existence of: **Non-linear** relationship **Outliers** **More than one** relationship	B1		Any one; OE Not reasons identifiable from context (eg spurious)	
	Sensible related sketch	B1	2		
	SC: Check on calculation $\Rightarrow$ B1 B0				
	Total		7		

Question 2: Jan 2011

1(a)(i)	$r = 0.6$ to 0.98	B1		AWFW (≈ 0.8) If answers are not labelled, assume order is (a)(i) then (a)(ii)	
(ii)	$r = -0.5$ to -0.02	B1	2	AWFW (≈ -0.3)	
	Accept answers as ranges if and only if contained entirely within given ranges			Eg: $\quad$ (a)(i) 0.7 to 0.9 $\Rightarrow$ B1 $\qquad$ (a)(ii) -0.6 to $-0.4 \Rightarrow$ B0	
(b)(i)	$r = 0.757$ $r = 0.75$ to 0.77 $r = 0.65$ to 0.85	B3 (B2) (B1)	3	AWRT (0.75708) AWFW AWFW	
	or				
	Attempt at $\sum x \ \sum x^2 \ \sum y \ \sum y^2$ and $\sum xy$	(M1)		271.5 6142.97 1911.9 304650.01 and 43259.17 (all 5 attempted)	
	or				
	Attempt at $S_{xx} \ S_{yy}$ and S_{xy}			0.2825 36.5425 and 2.4325 (all 3 attempted)	
	Attempt at substitution into correct corresponding formula for r $r = 0.757$	(m1) (A1)		AWRT	
(ii)	Strong/fairly strong/moderate positive (linear) correlation/relationship/ association/link (but not 'trend')	Bdep1		Dependent on $0.65 < r < 0.85$ Or equivalent; must qualify strength and indicate positive Bdep0 for very strong/high/average/ medium/some etc.	
	between				
	Circumference/size and weight of (cricket) balls	B1	2	Context; providing $0 < r < 1$	
	Total		7		

2(a)	$r = 0.893$ to 0.8933	B3		AWFW	(0.89319)
	$r = 0.89$ to 0.896	(B2)		AWFW	
	$r = 0.8$ to 0.95	(B1)		AWFW	
	or				
	Attempt at $\sum x$ $\sum x^2$ $\sum y$ $\sum y^2$ & $\sum xy$			561 30667 671 42613 & **35882** (all 5 attempted)	
	or	(M1)			
	Attempt at S_{xx} S_{yy} & S_{xy}			2056 1682 & **1661** (all 3 attempted)	
	Attempt at **correct** corresponding formula for r	(m1)			
	$r = 0.893$ to 0.8933	(A1)	3	AWFW	
(b)	Fairly strong / strong / very strong positive (linear) correlation / relationship / association / link (but not trend)	B1dep		Or equivalent; must qualify strength and indicate positive Dependant on $0.8 \leq r \leq 0.95$ B0 for some/average/medium/etc	
	between				
	length and weight of adult snakes	B1	2	Context; providing $0 < r < 1$	
(c)	Figure 1: 5 correct labelled points 4 or 3 correct labelled points	B2 (B1)	2	Deduct 1 mark if points not labelled	
(d)(i)	D and G	B1	1	Both CAO	
(ii)	$r = 0.25$ to 0.75	B1		AWFW (0.48790) No penalty for calculation Accept a range only if whole of it falls within 0.25 to 0.75	
	Fairly weak / weak / some / moderate positive (linear) correlation / relationship / association / link Do not accept comparison with value in (a) or statement in (b)	B1dep	2	Or equivalent; must qualify strength and indicate positive Dependant on $0.25 \leq r \leq 0.75$ B0 for very weak/little/slight/hardly any/fair/average/medium/anything involving strong/etc	
		Total	**10**		

2 (a)(i)	$r = 0.022$ to 0.023	B3		AWFW (0.022557)	
	$r = 0.02$ to 0.03	(B2)		AWFW	
	$r = -0.1$ to 0.1	(B1)		AWFW	
	OR				
	Attempt at $\sum x$ $\sum x^2$ $\sum y$ $\sum y^2$ & $\sum xy$			118.8 1619.36 31.5 114.43 & 416.13 (all 5 attempted)	
	or	(M1)			
	Attempt at S_{xx} S_{yy} & S_{xy}			51.2 4.18 & 0.33 (all 3 attempted)	
	Attempt at **correct** formula for r	(m1)			
	$r = 0.022$ to 0.023	(A1)	3	AWFW	
(ii)	(Almost/virtually) **no/zero** (linear) **correlation** (relationship/association/link)	B1		Or equivalent qualification of **NO** strength; do not follow-through from (i) B0 for very weak/weak/some/ little/slight/positive/hardly any/etc unless correct qualification also stated	
	between				
	length and (maximum) **diameter** of carrots	B1	2	Context; providing $-1 < r < 1$	
(b)	**Unlikely**/wrong/incorrect/invalid	B1		Or equivalent	
	Would expect a **positive value** **or** Would expect **weight to increase with length** **or** Would imply **shorter carrots are heavier**	B1	2	Or equivalent reason	
		Total	**7**		

Question 5: Jun 2007

1(a)	$r = -0.526$ to -0.525 or	B3		AWFW	
	$r = -0.53$ to -0.52 or	(B2)		AWFW; ignore sign	
	$r = -0.6$ to -0.4	(B1)		AWFW; ignore sign	
	OR				
	Attempt at $\sum x,\ \sum x^2,\ \sum y,\ \sum y^2$ and $\sum xy$ or	(M1)		260, 6970, 143, 2083 and 3671	
	Attempt at $S_{xx},\ S_{yy}$ and S_{xy}			210, 38.1 and -47	
	Attempt at a correct formula for r	(m1)			
	$r = -0.526$ to -0.525	(A1)	3	AWFW	
(b)				OE; must qualify strength and indicate negative	
	Weak/some/moderate negative correlation (relationship/association)	B1		B0 for strong/poor/reasonable/average B0 if $r > 0$ or $r < -1$ B0 if contradictory statements	
	between				
	length and (maximum) diameter	B1		Context	
	Ignore subsequent comments (as below) only if B1 B1 already scored				
	OR				
	Some evidence that large lengths are associated with small diameters	(B1) (B1)		OE; must qualify strength and indicate negative	
	OR				
	Longer melons tend to have smaller diameters / be thinner	(B1) (B1)	2	OE; must qualify strength and indicate negative	
	Total		**5**		

Question 6: Jun 2006

1(a)(i)	$r = 0.143$ to 0.1432 or	B3		AWFW	
	$r = 0.142$ to 0.144 or	B2		AWFW	
	$r = 0.1$ to 0.2	B1		AWRT	
	Attempt at $\quad\quad \Sigma x\ \ \Sigma x^2$ $\quad\quad\quad\quad\quad \Sigma y\ \ \Sigma y^2$ $\quad\quad\quad\quad\quad \Sigma xy$ or			3952, 2228282 47.00, 292.0000 23517.50	
	Attempt at $\quad\quad S_{xx}\ \ S_{yy}\ \ S_{xy}$	M1		275994, 15.875, 299.5	
	Attempt at a correct formula for r	m1			
	$r = 0.143$ to 0.1432	A1	3	AWFW	
(ii)	Little/weak/no correlation/relationship/association between number of pages and (retail)	B1		or equivalent; but not poor	
	price	B1	2	context	
(iii)	Size (page, thickness), author, ranking, publicity/marketing, cover design, recommendations on back, publisher, font, popularity, quality, print-run, etc	B1	1	or any sensible variable but not pictures, coloured pictures, age, words, weight, mass	
(b)	(Very) strong/almost exact positive/perfect correlation/relationship/ association	B1		or equivalent	
	between				
	number of pages and sale/new price	B1	2	context	
	Sale price appears to be determined by number of pages	B2		or equivalent	
	Total		**8**		

7					
(a)	**Ryan:** Value indicates that as **volume increases** then **weight decreases**	B1		Or equivalent in context	
	Sunil: Value indicates **no correlation/relationship/ association/link** between **volume and weight**	B1	2	Or equivalent in context	
	SC: If B0 B0: Would expect **weight** to **increase** with **volume** or Would expect **strong(er) positive** correlation between **weight** and **volume**	(B1)		Or equivalent in context	
(b)	**Ryan & Sunil:** r is not affected by units/(linear) scaling	B1		Or equivalent	
	Tim: r is not affected by sample size or $2 \times 0.612 > 1 \implies$ impossibility	B1	2	Either; or equivalent	
(c) (i)	$r = 0.541$ to 0.543 $r = 0.54$ to 0.55 $r = 0.5$ to 0.6	B3 (B2) (B1)	3	AWFW (0.54186) AWFW AWFW	
	OR				
	Attempt at $\sum v$ $\sum v^2$ $\sum w$ $\sum w^2$ & $\sum vw$ or Attempt at S_{vv} S_{ww} & S_{vw}	(M1)		216 6633.16 136 2376.84 & **3795.5** (all 5 attempted) Accept notation of x and y 801.16 64.84 & **123.5** (all 3 attempted)	
	Attempt at substitution into **correct** corresponding formula for r $r = 0.541$ to 0.543	(m1) (A1)		AWFW	
(ii)	(Quite or fairly) **weak/some/moderate** **positive** (linear) **correlation**/relationship/ association/link (*but not 'trend'*) between **volumes** and **weights** of suitcases	Bdep1 B1	 2	Dependent on $0.5 \le r \le 0.6$ Or equivalent; must **qualify strength** and **state positive** Bdep0 for very strong/strong/high/ good/average/medium/reasonable/ poor/very weak/little/etc Context; providing $0 < r < 1$	
	Total		9		