



eClassroom

INTERNATIONAL A LEVEL MATHEMATICS • STATISTICS 1

Statistics 1

Exam Questions by Topic

Pearson Edexcel International Advanced Level (WST01/01) • Calculator and Statistical Tables allowed

Multi-step exam questions at the difficulty of the real WST01 paper, with a full mark scheme

Chapters (tap a chapter to jump to it)	Qs	Marks	Page	My score
1 Mathematical Modelling Not examined on its own: modelling ideas are tested within Chapters 2 to 7	—	—	Not applicable	
2 Measures of Location and Spread Averages, quartiles, interpolation, standard deviation, coding	10	83	own PDF	
3 Representations of Data Histograms, stem and leaf, box plots, outliers, skewness, comparisons	10	80	own PDF	
4 Probability Venn and tree diagrams, conditional probability, independence	10	80	own PDF	
5 Correlation and Regression PMCC, least squares regression, interpretation, coding	10	84	2	
6 Discrete Random Variables Probability and cumulative distributions, $E(X)$, $\text{Var}(X)$, uniform	10	82	own PDF	
7 The Normal Distribution Standardising, tables, finding μ and σ , conditional problems	10	83	own PDF	

[Full mark scheme](#)

[Free courses](#)

[More worksheets](#)

Name

Class

Date



Join students already acing their exams.

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

CHAPTER 5 • KEY SKILLS

5

Correlation and Regression

Specification S1: correlation and regression • 10 questions • 84 marks • about 100 minutes

1 Scatter diagrams

Describe the direction and strength of the correlation, and name both variables in context.

2 Summary statistics

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n},$$

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$$

3 The PMCC

$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$, with $-1 \leq r \leq 1$. Give r to at least 3 s.f.

4 Interpreting r

r measures *linear* correlation only. Correlation does not show that one variable causes the other.

5 The regression line

$$y = a + bx \text{ with } b = \frac{S_{xy}}{S_{xx}} \text{ and } a = \bar{y} - b\bar{x}.$$

Write the final equation.

6 Explanatory variable

The explanatory variable (x) is set or controlled; the response (y) depends on it. Only predict y from x .

7 Interpreting a and b

b : the change in y for each increase of 1 in x , in context and units. a may be meaningless if $x = 0$ is impossible.

8 Predictions

Within the data (interpolation) is reliable; outside it (extrapolation) is not. Check that the prediction is possible.

9 Coding

Coding does not change r . Substitute the codes into the coded line to get the line in the original variables.

Skills map every skill above is tested at least once; interleaved skills come from GCSE or other chapters

Skill	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Marks	5	5	5	5	5	13	11	13	10	12
1 Scatter diagrams			•							
2 Summary statistics	•					•		•		•
3 The PMCC	•			•		•	•	•	•	•
4 Interpreting r			•			•			•	•
5 The regression line		•				•	•	•	•	•
6 Explanatory variable			•						•	
7 Interpreting a and b				•	•	•	•		•	
8 Predictions		•			•	•	•	•		
9 Coding				•				•		
Interleaved: coding and correcting data (Chapter 2)				•				•		•
Interleaved: straight lines, rearranging (P1)				•				•	•	
Interleaved: units and conversions					•		•			

Exam difficulty. Questions 1 to 5 check each skill on its own. Questions 6 to 10 are multi-part questions built like the 11 to 14 mark questions on a real WST01 paper, where one part leads into the next and skills are combined.

Must learn. Every interpretation must name the variables and use their units. In “show that” questions give values to more accuracy than the answer you are shown.



Join students already acing their exams.

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

1. For 10 pairs of observations (x, y) :

$$n = 10 \quad \sum x = 55 \quad \sum y = 320 \quad \sum x^2 = 385 \quad \sum y^2 = 10\,650 \quad \sum xy = 1920$$

- (a) Find S_{xx} , S_{yy} and S_{xy}

(3)

- (b) Calculate the product moment correlation coefficient between x and y .

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 1 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 1 is 5 marks)**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

2. For a set of data with x values between 2 and 10:

$$\bar{x} = 6 \quad \bar{y} = 20 \quad S_{xx} = 40 \quad S_{xy} = 96$$

- (a) Find the equation of the regression line of y on x in the form $y = a + bx$

(3)

- (b) Estimate the value of y when $x = 7$

(1)

- (c) Give a reason why an estimate of y when $x = 25$ may be unreliable.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 2 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 2 is 5 marks)**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

3.

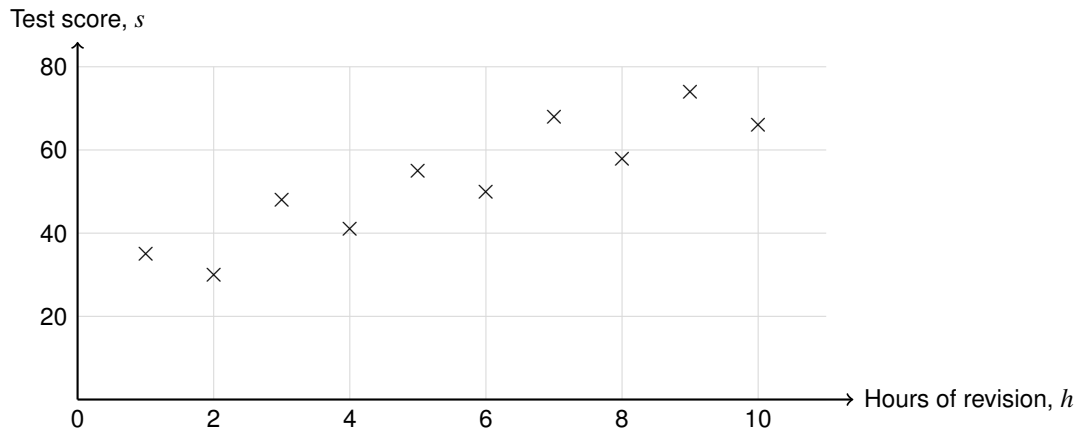


Figure 1

Figure 1 shows the number of hours of revision, h , and the test score, s , for 10 students. The product moment correlation coefficient for these data is 0.902

- (a) Interpret this value in the context of the question. (1)
- (b) State which variable is the explanatory variable, giving a reason. (1)
- (c) A teacher says: “This shows that revising for longer causes higher test scores.” Comment on this statement. (1)
- For a different group, the product moment correlation coefficient between h and s is -0.12
- (d) Interpret this value. (1)
- (e) State one limitation of the product moment correlation coefficient as a measure of the relationship between two variables. (1)



Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 3 is 5 marks)**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

4. Data on two variables x and y are coded using $u = x - 100$ and $v = \frac{y}{10}$

The product moment correlation coefficient between u and v is 0.72, and the regression line of v on u is $v = 2.1 + 0.35u$

- (a) Write down the product moment correlation coefficient between x and y . (1)
- (b) Find the equation of the regression line of y on x in the form $y = a + bx$ (3)
- (c) Interpret the value of b in your line. (1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 4 is 5 marks)**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

- (a) Interpret the value 0.156 in context. (2)
- (b) Estimate, in dollars, the monthly amount spent on stationery by an office with 80 employees. (2)
- (c) Explain why the value 12.5 should not be interpreted as the amount spent by an office with no employees. (1)

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 5 is 5 marks)



Join students already acing their exams.

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

6. A café records the midday temperature, t °C, and the number of ice creams sold, g tens, on 8 days.

t (°C)	14	16	18	20	22	24	26	28
g (tens)	20	25	22	30	29	35	33	41

(You may use $\sum t = 168$, $\sum g = 235$, $\sum t^2 = 3696$, $\sum g^2 = 7245$, $\sum tg = 5160$)

- (a) Show that $S_{tg} = 225$ and find the values of S_{tt} and S_{gg} (3)
- (b) Calculate the product moment correlation coefficient between t and g . (2)
- (c) Interpret your value in context. (1)
- (d) Find the equation of the regression line of g on t in the form $g = a + bt$ (4)
- (e) Interpret the value of b in context. (1)
- (f) (i) Estimate the number of ice creams sold on a day when the temperature is 25 °C. (1)
- (ii) Comment on the reliability of using the line to estimate sales on a day when the temperature is 40 °C. (1)



Question 6 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

Question 6 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 6 is 13 marks)



Join students already acing their exams.

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

7. A biologist measures the length, L cm, and the mass, m grams, of 9 fish. The lengths are between 22 cm and 48 cm. The data are summarised by

$$\sum L = 315 \quad \sum m = 3880 \quad S_{LL} = 586 \quad S_{mm} = 234\,889 \quad S_{Lm} = 11\,670$$

- (a) Calculate the product moment correlation coefficient between L and m . (2)
- (b) Show that the regression line of m on L is $m = -266 + 19.9L$, with coefficients given to 3 significant figures. (4)
- (c) Estimate the mass of a fish of length 0.36 m. (2)
- (d) Explain why the regression line should not be used to estimate the mass of a fish of length 10 cm. (2)
- (e) Interpret the value 19.9 in context. (1)



Question 7 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

Question 7 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 7 is 11 marks)



Join students already acing their exams.

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

8. The price, £ p , of a product in year t was recorded every 4 years from 2000 to 2020. The data are coded using

$$x = \frac{t - 2000}{4} \quad y = \frac{p - 200}{10}$$

and summarised by

$$n = 6 \quad \sum x = 15 \quad \sum y = 41 \quad \sum x^2 = 55 \quad \sum y^2 = 371.9 \quad \sum xy = 141.8$$

- (a) Find S_{xx} , S_{yy} and S_{xy} (3)
- (b) Calculate the product moment correlation coefficient between x and y . (2)
- (c) Write down the product moment correlation coefficient between t and p , giving a reason. (1)
- (d) Find the equation of the regression line of y on x . (3)
- (e) Hence find the equation of the regression line of p on t in the form $p = c + d(t - 2000)$ (2)
- (f) Use your line to estimate the price in 2030, and comment on the reliability of your estimate. (2)



Question 8 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

Question 8 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 8 is 13 marks)



Join students already acing their exams.

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

- $$\Sigma a = 540 \quad \Sigma s = 1560 \quad S_{aa} = 2250 \quad S_{ss} = 1200 \quad S_{as} = 1350$$

- (a) Calculate the product moment correlation coefficient between a and s . (2)
- (b) Find the equation of the regression line of s on a . (3)
- (c) Interpret the gradient of your line in context. (1)
- (d) Using your line, find the increase in age that corresponds to an increase in blood pressure of 9 mmHg. (2)
- (e) Explain why the line should not be used to estimate the age of a patient from their blood pressure. (1)
- (f) The doctor says: "The value of r proves that getting older causes blood pressure to rise." Comment on this statement. (1)

Question 9 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

Question 9 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 9 is 10 marks)**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

$$\Sigma x = 60 \quad \Sigma y = 350 \quad \Sigma x^2 = 440 \quad \Sigma y^2 = 13050 \quad \Sigma xy = 2280$$

- The student then found that the pair recorded as $(8, 30)$ should have been $(8, 45)$

- (c) Calculate the correct product moment correlation coefficient, and describe the effect of the correction. (3)

- (d)** Find the equation of the correct regression line of y on x . **(3)**

Question 10 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk

Question 10 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 10 is 12 marks)**TOTAL FOR CHAPTER 5 IS 84 MARKS****Join students already acing their exams.**Book your FREE lesson at www.eclassroom.org.uk