A Level Statistics
AQA Past Exam Questions

TOPIC: Least Squares Regression

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have retrievable mathematical formulae stored in them.

Instructions

• Use **black** ink or ball-point pen.

• If pencil is used for diagrams/sketches/graphs it must be dark (HB or B). Coloured pencils and highlighter pens must not be used.

• **Fill in the boxes** at the top of this page with your name, centre number and candidate number.

• Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.

• Answer the questions ***on paper***

• You should show sufficient working to make your methods clear. Answers without working may not gain full credit.

• Unless otherwise stated, statistical tests should be carried out at the 5% significance level.

• When a calculator is used, the answer should be given to three significant figures unless otherwise stated.

Information

• ***You may use the*** booklet 'Statistical Formulae and Tables'

• There are **12** questions in this question paper. The total mark for this paper is ***123***

• The marks for **each** question are shown in brackets – use this as a guide as to how much time to spend on each question.

Advice

• Read each question carefully before you start to answer it.

• Try to answer every question.

• Check your answers if you have time at the end.

• If you change your mind about an answer, cross it out and put your new answer and any working underneath.

• Check your answers if you have time at the end.

AQA_JAN_2012_5

An experiment was undertaken to collect information on the burning of a specific type of wood as a source of energy. At given fixed levels of the wood's moisture content, x per cent, its corresponding calorific value, y MWh=tonne, on burning was determined. The results are shown in the table.

x	5	10	15	20	25	30	35	40	45	50	55	60	65
y	5.2	4.7	4.3	4.0	3.2	2.8	2.5	2.2	1.8	1.5	1.3	1.0	0.6

(a) Explain why calorific value is the response variable.

(1 mark)

(b) Calculate the equation of the least squares regression line of y on x, giving your answer in the form y = a + bx

(5 marks)

(c) Interpret, in context, your values for a and b.

(3 marks)

(d) Use your equation to estimate the wood's calorific value when it has a moisture content of 27 per cent.

(2 marks)

(e) Calculate the value of the residual for the point (35, 2.5).

(2 marks)

(f) Given that the values of the 13 residuals lie between -0.28 and +0.23 , comment on the likely accuracy of your estimate in part (d).

(1 mark)

(g) (i) Give a general reason why your equation should not be used to estimate the wood's calorific value when it has a moisture content of 80 per cent.

(1 mark)

(ii) Give a specific reason, based on the context of this question and with numerical support, why your equation cannot be used to estimate the wood's calorific value when it has a moisture content of 80 per cent.

(2 marks)

AQA_JAN_2013_1

Bob, a church warden, decides to investigate the lifetime of a particular manufacturer's brand of beeswax candle. Each candle is 30 cm in length.

From a box containing a large number of such candles, he selects one candle at random. He lights the candle and, after it has burned continuously for x hours, he records its length, y cm, to the nearest centimetre. His results are shown in the table.

x	5	10	15	20	25	30	35	40	45
y	27	25	21	19	16	11	9	5	2

(a) State the value that you would expect for a in the equation of the least squares regression line, y = a + bx

(1 mark)

(b) (i) Calculate the equation of the least squares regression line, y = a + bx

(4 marks)

(ii) Interpret the value that you obtain for b.

(2 marks)

(iii) It is claimed by the candle manufacturer that the total length of time that such candles are likely to burn for is more than 50 hours. Comment on this claim, giving a numerical justification for your answer.

(2 marks)

AQA_JAN_2010_3

The table shows, for each of a random sample of 7 weeks, the number of customers, x, who purchased fuel from a filling station, together with the total volume, y litres, of fuel purchased by these customers.

x	230	184	165	147	241	174	210
y	4551	3410	3252	3756	3787	4024	4254

(a) Calculate the equation of the least squares regression line of y on x.

(4 marks)

(b) Estimate the volume of fuel sold during a week in which 200 customers purchase fuel.

(2 marks)

(c) Comment on the likely reliability of your estimate in part (b), given that, for the regression line calculated in part (a), the values of the 7 residuals lie between approximately -415 litres and +430 litres.

(2 marks)

AQA_JAN_2011_5

Craig uses his car to travel regularly from his home to the area hospital for treatment. He leaves home at x minutes after 7.30 am and then takes y minutes to arrive at the hospital's reception desk.

His results for 11 mornings are shown in the table.

x	0	5	10	15	20	25	30	35	40	45	50
y	31	42	32	58	47	56	79	68	89	95	85

(a) Explain why the time taken by Craig between leaving home and arriving at the hospital's reception desk is the response variable.

(1 mark)

(b) Calculate the equation of the least squares regression line of y on x, writing your answer in the form $y = a + bx$.

(5 marks)

(c) On a particular day, Craig needs to arrive at the hospital's reception desk no later than 9.00 am. He leaves home at 7.45 am.

Estimate the number of minutes before 9.00 am that Craig will arrive at the hospital's reception desk. Give your answer to the nearest minute.

(5 marks)

(d) (i) Use your equation to estimate y when $x = 85$.

(1 mark)

(ii) Give one statistical reason and one reason based on the context of this question as to why your estimate in part (d)(i) is unlikely to be realistic.

(2 marks)

AQA_JUNE_2008_1

The table shows the times taken, y minutes, for a wood glue to dry at different air temperatures, x°C.

x	10	12	15	18	20	22	25	28	30
y	42.9	40.6	38.5	35.4	33.0	30.7	28.0	25.3	22.6

(a) Calculate the equation of the least squares regression line $y = a + bx$

(4 marks)

(b) Estimate the time taken for the glue to dry when the air temperature is 21°C.

(2 marks)

AQA_JUNE_2007_5

Bob, a gardener, measures the time taken, y minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, x ^{o}C. His results are shown in the table.

x	16	20	24	28	32	36	40	44	48	52	56
y	4.7	4.3	3.8	3.5	3.0	2.7	2.4	2.0	1.8	1.6	1.1

(a) State why the explanatory variable is temperature.

(1 mark)

(b) Calculate the equation of the least squares regression line y = a + bx

(4 marks)

(c) (i) Interpret, in the context of this question, your value for b.

(2 marks)

(ii) Explain why no sensible practical interpretation can be given for your value of a.

(2 marks)

(d) (i) Estimate the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 30^{o}C.

(2 marks)

(ii) Show why the equation cannot be used to make a valid estimate of the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 75 ^{o}C.

(2 marks)

AQA_JUNE_2009_4

As part of an investigation, a chlorine block is immersed in a large tank of water held at a constant temperature. The block slowly dissolves, and its weight, y grams, is noted x days after immersion. The results are shown in the table.

x days	5	10	15	20	30	40	50	60	75
y grams	47	44	42	38	35	27	23	16	9

(a) Calculate the equation of the least squares regression line of y on x.

(4 marks)

(b) Hence estimate, to the nearest gram, the initial weight of the block.

(1 mark)

(c) A company which markets the chlorine blocks claims that a block will usually dissolve completely after about 13 weeks. Comment, with justification, on this claim.

(3 marks)

AQA_JUNE_2011_3a

During a particular summer holiday, Rick worked in a fish and chip shop at a seaside resort.
He suspected that the shop's takings, £y, on a weekday were dependent upon the forecast of that day's maximum temperature, x°C, in the resort, made at 6.00 pm on the previous day.
To investigate this suspicion, he recorded values of x and y for a random sample of 7 weekdays during July.

x	23	18	27	19	25	20	22
y	4290	3188	5106	3829	5057	4264	4485

(i) Calculate the equation of the least squares regression line of y on x.

(4 marks)

(ii) Estimate the shop's takings on a weekday during July when the maximum temperature was forecast to be 24°C.

(2 marks)

(iii) Explain why your equation may not be suitable for estimating the shop's takings on a weekday during February.

(1 mark)

(iv) Describe, in the context of this question, a variable other than the maximum temperature, x, that may affect y.

(1 mark)

AQA_JUNE_2014_3

The table shows the body mass index (BMI), x , and the systolic blood pressure (SBP), y mmHg, for each of a random sample of 10 men, aged between 35 years and 40 years, from a particular population.

x	13	23	29	35	17	34	25	20	31	27
y	103	115	124	126	108	120	113	117	118	119

(a) Calculate the equation of the least squares regression line of y on x .

[4 marks]

(b) Use your equation to estimate the SBP of a man from this population who is aged 38 years and who has a BMI of 30 .

[2 marks]

(c) State why your equation might not be appropriate for estimating the SBP of a man from this population:
(i) who is aged 38 years and who has a BMI of 45 ;
(ii) who is aged 50 years and who has a BMI of 25 .

[2 marks]

(d) Find the value of the residual for the point (20, 117).

[2 marks]

(e) The mean of the vertical distances of the 10 points from the regression line calculated in part (a) is 2.71, correct to three significant figures. Comment on the likely accuracy of your estimate in part (b).

[1 mark]

AQA_JUNE_2016_5

The table shows the ground temperature, x°C, at 6 pm and the number of wing vibrations per second, y, made by a particular type of insect, called a striped ground cricket.

x	31.4	25.0	34.1	30.7	27.1	24.3	20.6	21.8	28.9	26.9	28.6	24.6
y	19.0	16.0	19.8	18.0	17.3	15.5	14.7	15.4	18.1	16.8	17.0	16.4

(a) (i) Calculate the equation of the least squares regression line in the form y a + bx

[4 marks]

(ii) Interpret your value for b in the context of the question.

[2 marks]

(iii) Given that the wings of striped ground crickets do not vibrate at temperatures below 15°C, explain why your value for a has no practical interpretation.

[2 marks]

(b) Estimate the number of wing vibrations of a striped ground cricket when the ground temperature at 6 pm is 23°C.

[1 mark]

(c) (i) Calculate the value of the residual for the point (28.6 , 17.0).

[2 marks]

(ii) State why the sum of the residuals for the 12 points gives no useful information about the strength of the linear relationship between y and x .

[1 mark]

AQA_JUNE_2017_2

Ten women, aged between 45 years and 65 years and who have previous experience of browsing the internet, were asked to record their total times spent browsing the internet during a period of seven days. The women's ages, x years, and their recorded times, y hours, are shown in the table.

x	46.8	50.0	51.2	55.3	56.4	59.3	61.0	62.2	63.0	64.8
y	22.4	23.1	20.6	21.8	19.3	17.0	15.3	16.6	12.7	13.2

(a) Calculate the equation of the least squares regression line, giving your answer in the form y = a + bx .

[5 marks]

(b) The 10 women can be regarded as a random sample from women aged between 45 years and 65 years who have previous experience of browsing the internet.

(i) Estimate the total time spent browsing the internet during a period of seven days by a woman aged 60.0 years who has previous experience of browsing the internet.

[1 mark]

(ii) Give two reasons why your equation, calculated in part (a), may not be appropriate for estimating the total time spent browsing the internet during a period of seven days by a randomly selected woman.

[2 marks]

AQA_JUNE_2018_5

Ethylene glycol, commonly known as antifreeze, is added to water in a vehicle's radiator to lower the freezing point of the resultant fluid.

The table shows the percentage, x , of antifreeze by volume in the fluid and the fluid's corresponding freezing point, y ºC.

x	0.0	5.0	10.0	15.0	20.0	25.0	30.0
y	0.0	–2.1	– 4.1	–6.3	–8.3	–11.9	–14.7

(a) State why the least squares regression line of x on y would not be appropriate for these data.

[1 mark]

(b) (i) Calculate the equation of the least squares regression line of y on x .

[3 marks]

(ii) Hence interpret, in context, the value you obtained for the line's gradient.

[2 marks]

(c) (i) Calculate the value of the residual for the point (15, –6.3).

[2 marks]

(ii) Hence, given that the residual for the point (30, –14.7) is –0.65 , correct to two significant figures, find the sum of the remaining 5 residuals.

[2 marks]