A-LEVEL MATHS (EDEXCEL 9MA0) · STATISTICS YEAR 2 · CHAPTER 1: REGRESSION, CORRELATION AND HYPOTHESIS TESTING · A* EXAM PRACTICE

Hints online, answers free with an account:
mathswariz.uk/worksheets/a-level-statistics-year-2-chapter-1-a-star/

Regression, correlation and hypothesis testing: A* questions

78 marks

Date

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X A* exam questions

A to A*

Extended Edexcel-style questions. Show every step. The hints are at the end of the sheet. Use one at a time.

X1 An engineer records values of x and y for $1 \leqslant x \leqslant 8$. She codes the data with $p = \log x$ and $q = \log\left(\frac{y}{3}\right)$, and finds the regression line of q on p to be $q = 0.215 + 1.48p$. Show that the data can be modelled by $y = ax^n$, giving a and n to 3 significant figures. Use your model to estimate the value of x for which $y = 50$, and comment on the reliability of this estimate. [8 marks]

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X2 A metal casting is taken out of a furnace and left to cool in a workshop, which is kept at 20 °C. Its temperature, T °C, is measured x minutes later. An engineer suggests the model $T = 20 + kb^x$, where k and b are constants.

x (minutes)	T (°C)
0	330
4	207
8	131
12	88
16	60
20	44
24	34
28	29

(a) Find the PMCC between x and $\log T$, and the PMCC between x and $\log(T - 20)$. [3 marks]

Answer .....

(b) Find the regression line of $\log(T - 20)$ on x. Hence find k and b, giving each to 3 significant figures. [4 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) The casting can be handled safely when its temperature is below 100 °C. Use your model to find how long after it leaves the furnace this is. [3 marks]

Answer .....

(d) Using your answers to part (a), and what happens to T as x becomes large, explain why the model $T = 20 + kb^x$ is more suitable than a model of the form $T = kb^x$. [3 marks]

(e) Interpret the value of b in context. [2 marks]

Total for X2: 15 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X3 For a random sample of 26 students, the PMCC between the time spent on social media and the hours of sleep is r. Asha tests, at the 1% level, whether there is negative correlation, and rejects H_0. Ben uses the same data to test, at the 1% level, whether there is any correlation, and does not reject H_0. Find the range of possible values of r, and explain how the two tests can give different results. [8 marks]

Sample size n	0.05	0.025	0.01	0.005
25	0.3365	0.3961	0.4622	0.5052
26	0.3297	0.3882	0.4534	0.4958
27	0.3233	0.3809	0.4451	0.4869

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X4 The leaf area, $y\ \text{cm}^2$, of a plant is measured at ages x weeks, for $2 \leqslant x \leqslant 10$. Two codings are tried. The regression line of $\log y$ on x is $\log y = 0.48 + 0.105x$, with PMCC 0.962. The regression line of $\log y$ on $\log x$ is $\log y = 0.12 + 1.35 \log x$, with PMCC 0.995.

(a) Find the two models, in the forms $y = kb^x$ and $y = ax^n$, giving constants to 3 significant figures. [4 marks]

Answer .....

(b) State, with a reason, which model is more suitable, and use it to estimate the leaf area at 6 weeks. [3 marks]

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) Find what each model predicts for $x = 30$, and comment. [4 marks]

Total for X4: 11 marks

X5 A shop records the daily maximum temperature, x °C, and the number of ice creams sold, y, on 15 randomly chosen days. To make the numbers easier, a student codes the data with $u = 50 - x$ and $v = 2y$, and finds that the PMCC between u and v is -0.58.

(a) State the PMCC between x and y, giving a reason. [3 marks]

Sample size n	0.05	0.025	0.01
14	0.4575	0.5324	0.6120
15	0.4409	0.5140	0.5923
16	0.4259	0.4973	0.5742

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(b) Test, at the 2.5% level of significance, whether there is positive correlation between temperature and ice cream sales. [5 marks]

(c) Another student says, "If we use $w = \log y$ instead of y, the PMCC between x and w must also be 0.58." Explain whether he is right. [3 marks]

Total for X5: 11 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X6 A teacher believes that students who revise for longer get higher test scores. She collects data from two schools. In school B, for a random sample of 14 students, the PMCC between the hours of revision and the test score is 0.41. She then adds 12 students from school A, and for all 26 students together the PMCC is 0.47.

(a) Using the combined sample of 26 students, test the teacher's belief at the 5% level of significance. [6 marks]

Sample size n	0.05	0.025	0.01
14	0.4575	0.5324	0.6120
26	0.3297	0.3882	0.4534

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(b) Show that the data from school B on their own do not give evidence at the 5% level, and explain how the combined sample can give evidence when school B does not. [4 marks]

(c) Explain why the test in part (a) may not be valid. [3 marks]

Total for X6: 13 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X7 The table shows the mean distance from the Sun, R (in units where the Earth's distance is 1), and the time for one orbit, T years, of six planets.

R	**0.387**	**0.723**	**1**	**1.524**	**5.203**	**9.537**
T	0.241	0.615	1	1.881	11.86	29.46

(a) Using the coding $X = \log R$ and $Y = \log T$, show that the PMCC between X and Y is greater than 0.9999. [2 marks]

Answer .....

(b) Find the regression line of Y on X, and hence a model of the form $T = aR^n$, giving a and n to 3 significant figures. [4 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) Show that your model is the same as $T^2 = R^3$, and interpret a. [2 marks]

(d) Ceres orbits at a distance of 2.77. Estimate the time for one orbit of Ceres. [2 marks]

Answer .....

(e) A student wants to carry out a hypothesis test for positive correlation between X and Y. Explain why such a test is not appropriate here. [2 marks]

Total for X7: 12 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

F How did it go?

Colour one circle on each line.

	Not yet	Nearly	Yes
I can turn any log coding back into a model $y = ax^n$ or $y = kb^x$, including a coding with a constant in it.	○	○	○
I can say which model fits better, interpret its constants, and explain why two models can disagree outside the data.	○	○	○
I can find the range of r for which one test rejects and another does not.	○	○	○
I can judge whether a correlation test is valid for a given sample.	○	○	○

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

Hints

Use one hint at a time. Cover the ones below it with a sheet of paper.

X1	**Hint 1**	Write $\log\left(\frac{y}{3}\right) = \log y - \log 3$ and make $\log y$ the subject.
	Hint 2	If $\log y = c + n\log x$, then $y = 10^c x^n$.
	Hint 3	Is $y = 50$ reached at an x inside the data, or outside it?
X2	**Hint 1**	Make a column of $T - 20$ before you take logs.
	Hint 2	$\log(T-20) = \log k + x\log b$: the intercept is $\log k$ and the gradient is $\log b$.
	Hint 3	As x grows, b^x gets close to 0. What temperature does each model give then?
X3	**Hint 1**	Write the hypotheses for each test first. Which column of the table does each test use?
	Hint 2	Asha rejects: r is in her critical region. Ben does not reject: r is NOT in his.
	Hint 3	Put the two conditions together on a number line.
X4	**Hint 1**	From $\log y = c + mx$: $y = 10^c(10^m)^x$. From $\log y = c + n\log x$: $y = 10^c x^n$.
	Hint 2	A PMCC closer to 1 means the coded points are closer to a straight line.
	Hint 3	Compare the predictions at $x = 6$ with those at $x = 30$. Where were the data?
X5	**Hint 1**	Think of what each coding does to the order of the values and to their spacing.
	Hint 2	The hypotheses are about the population PMCC between x and y, so use the value of r you found in (a).
	Hint 3	A linear coding keeps the points in a straight line if they were in one. Does taking logs?

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

X6	**Hint 1**	Use all six steps for part (a); the sample size is the size of the combined sample.
	Hint 2	For school B you only need a new critical value, a comparison and a conclusion.
	Hint 3	What does the test assume about where the sample comes from?
X7	**Hint 1**	Make columns of $\log R$ and $\log T$ and use the statistics mode of your calculator.
	Hint 2	$\log T = \log a + n \log R$: the gradient is n and the intercept is $\log a$.
	Hint 3	What does a hypothesis test for correlation use the sample to learn about?

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.