



Name: _____

EXAM PAPERS PRACTICE

Edexcel GCSE Statistics

The collection of data(Higher)

Date:

Time:

Total marks available:

Total marks achieved: _____

© 2026 Exam Papers Practice. All Rights Reserved

Questions

Q1.

Julie was investigating the relationship between the marks gained by students in their GCSE Mathematics exam and the marks gained by the same students in an A-level Mathematics exam.

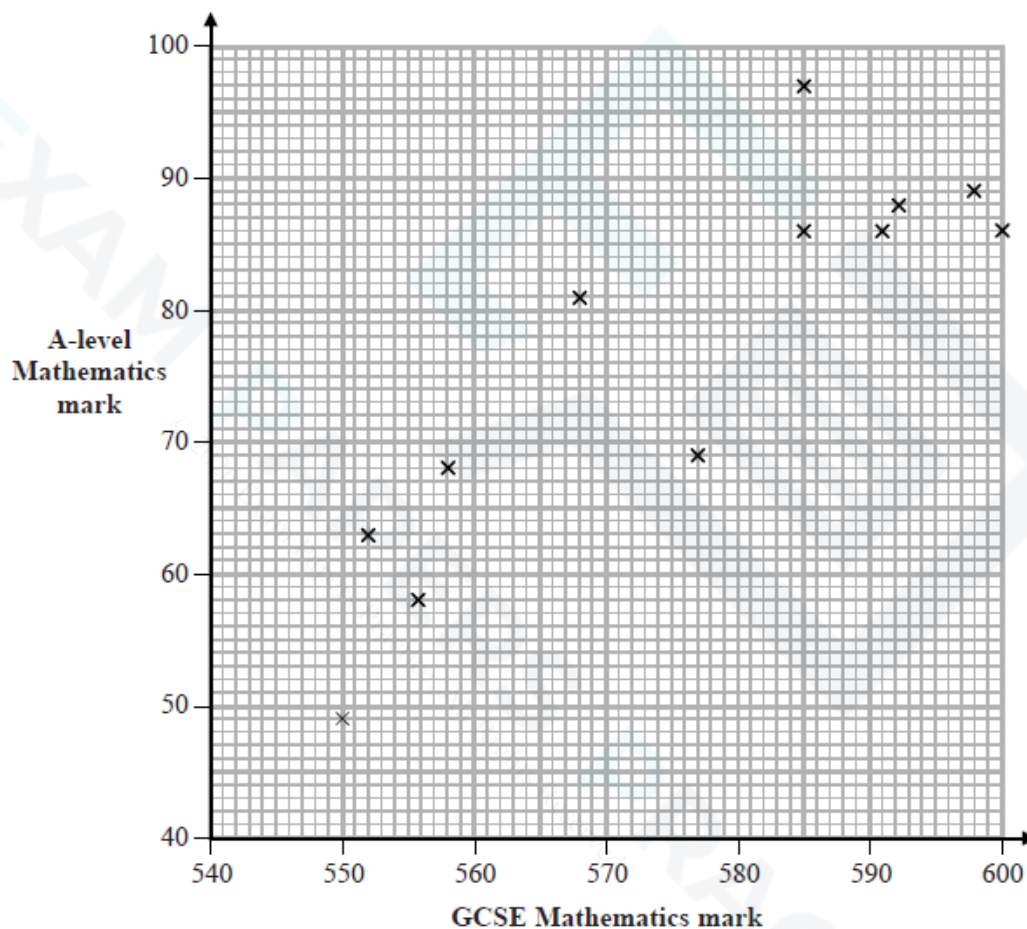
(a) Suggest a hypothesis Julie could use.

.....

.....

(1)

Julie drew a scatter diagram using the marks gained in GCSE Mathematics and the marks gained in A-level Mathematics by each of 12 students.



(b) For this scatter diagram, explain why the GCSE Mathematics mark is the explanatory variable.

.....

.....

.....

(1)

(c) Explain, giving a statistical reason, whether or not the scatter diagram supports your hypothesis in part (a).

.....



(2)

Using statistical software, Julie obtained the following information for her 12 students.

Mean GCSE Mathematics mark	578
Mean A-level Mathematics mark	78
Gradient of line of best fit	0.7

(d) Using this information, draw a line of best fit on the scatter diagram.

(2)

(e) Interpret the gradient of the line of best fit.

(1)

Julie wants to use the line of best fit to predict the A-Level Mathematics mark for a 13th student. For GCSE Mathematics this student gained a mark of 540

(f) Explain whether or not it would be appropriate to use the line of best fit to make her prediction.

(1)

(Total for question = 8 marks)

Q2.

A database contains information about the 2567 people registered at a medical centre.

Here is the information that the database contains on each person.

Name	Date of Registration	Age	Phone Number	Address	Postcode	Number of Appointments
------	----------------------	-----	--------------	---------	----------	------------------------

Rebecca wants to see if there is a relationship between the age of a person registered at the

medical centre and the number of appointments this person has had at the medical centre.

Describe how Rebecca can plan to use technology to help her process and present the data for this investigation.

You should refer to the advantages of using technology to process and present data compared to doing so by hand.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Total for question = 6 marks)

Q3.

Matthew is investigating how the cost of computer data storage has changed over time.

(a) Suggest a hypothesis Matthew could use.

.....

.....

Matthew finds the cost per terabyte of computer data storage in each first quarter of five successive years.

The table gives cost per terabyte, in US\$, for computer data storage. It also gives two of the chain base index numbers, correct to one decimal place, for this information.

Year	2015	2016	2017	2018	2019
Cost per terabyte (US\$)	33.33	31.25	29.00	24.87	24.43
Chain base index number		93.8	92.8		

(Source: *jcm.it.net*)

(b) Find the chain base index numbers for 2018 and 2019 and write them in the table.

Give each value correct to one decimal place.

(2)

(c) (i) Find the geometric mean of the four chain base index numbers.

You must show your working.
Give your answer correct to one decimal place.

.....

(2)

(ii) Interpret your answer.

.....

(2)

(Total for question = 7 marks)

Q4.

Irina is investigating whether the percentage of the population of a country living in urban areas has an effect on the life expectancy in that country.

(a) Suggest a hypothesis Irina could use for her investigation.

.....

.....

(1)

Irina collected the following information about 10 countries.

The Urban population (%) refers to the percentage of the population of the country who live in urban areas.

Country	Urban population (%)	Life expectancy (years)
Australia	89	82
Bangladesh	34	73
China	56	76
Egypt	43	73
France	80	82
Macedonia	57	76
Norway	81	82
Panama	67	79
Thailand	50	75
USA	82	80

(Source: CIA World Factbook)

Irina used statistical software to draw a scatter diagram for the information in the table.

(b) Give a reason why a scatter diagram is an appropriate diagram to use.

.....

.....

(1)

(c) For this investigation, which variable is the explanatory variable?
Give a reason for your answer.

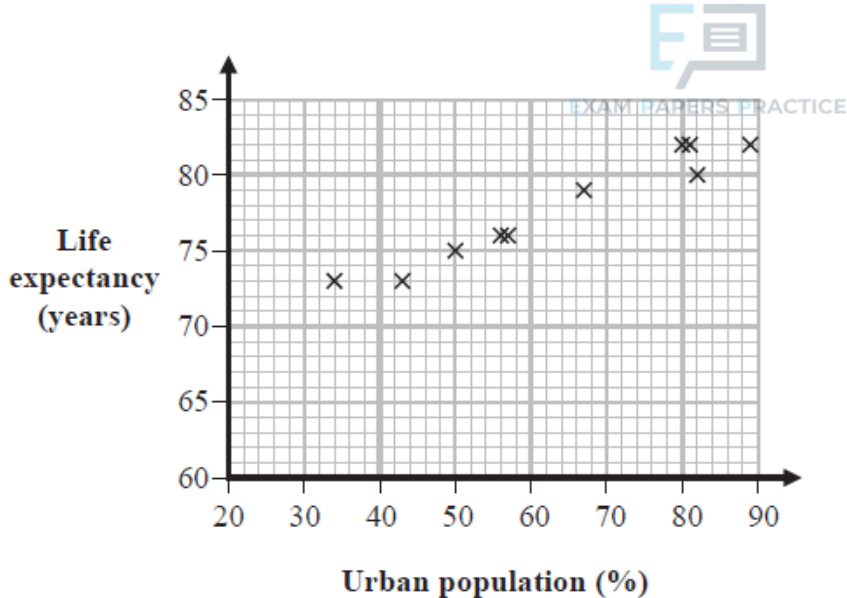
.....

.....

.....

(2)

The scatter diagram from the statistical software is shown below.



(d) Explain, giving a statistical reason, whether or not this scatter diagram supports your hypothesis in part (a).

.....

(2)

For these 10 countries, the double mean point of the data is (63.9, 77.8).

(e) Using this information, draw a line of best fit on the scatter diagram.

(2)

Using statistical software, Irina finds that the gradient of the line of best fit should be 0.19

(f) Interpret the gradient of the line of best fit.

.....

(1)

Irina now finds that South Africa has Urban population 65% and Life expectancy 63 years.

(g) Determine how this information for South Africa fits with the relationship shown in the scatter diagram for the other countries.

.....

(2)

(Total for question = 11 marks)

Q5.

Reeta read a report that said that these days more 18 to 24 year olds find out what is happening in the news from social media than find out by watching TV.

She decided to investigate how the 12 000 students at her university find out what is happening in the news.

Reeta wrote the following as a hypothesis:

Do more students get their news from social media than from other sources?

(a) Comment on whether it is appropriate to use this as a hypothesis.

.....

.....

(1)

Reeta planned to use a sample of the students at her university stratified by gender and by age.

(b) Explain why this method of sampling would be appropriate.

.....

.....

(1)

Reeta found that she could not get a list of all the students at her university to use for her stratified sample.

Instead she decided she would question students in the university cafeteria.

Here is the data collection sheet Reeta plans to use.

Access to news	Male	Female
Newspapers		
Online subscription news service		
Television news		
Social media		

(c) Discuss whether Reeta's data collection sheet is appropriate for her to use.

You should consider how Reeta might use the responses and describe any problems she may have when she uses the data collection sheet.

.....

.....



(3)

Reeta's friend George is at a different university with 18 000 students.

George carried out a similar investigation at his university so that he and Reeta could compare their results.

They each questioned the same percentage of students at the university they each attend.

They decided to compare their results by using comparative pie charts.

(d) Explain why comparative pie charts would be appropriate.

(1)

Reeta used a circle with diameter 10 cm for her pie chart.

(e) Calculate the diameter of the circle that George used for his pie chart.

..... cm

(2)

(f) Discuss **two** things that Reeta and George should have considered in planning their investigations to help improve the reliability and validity of their comparisons.

(2)

(Total for question = 7 marks)

Q6.

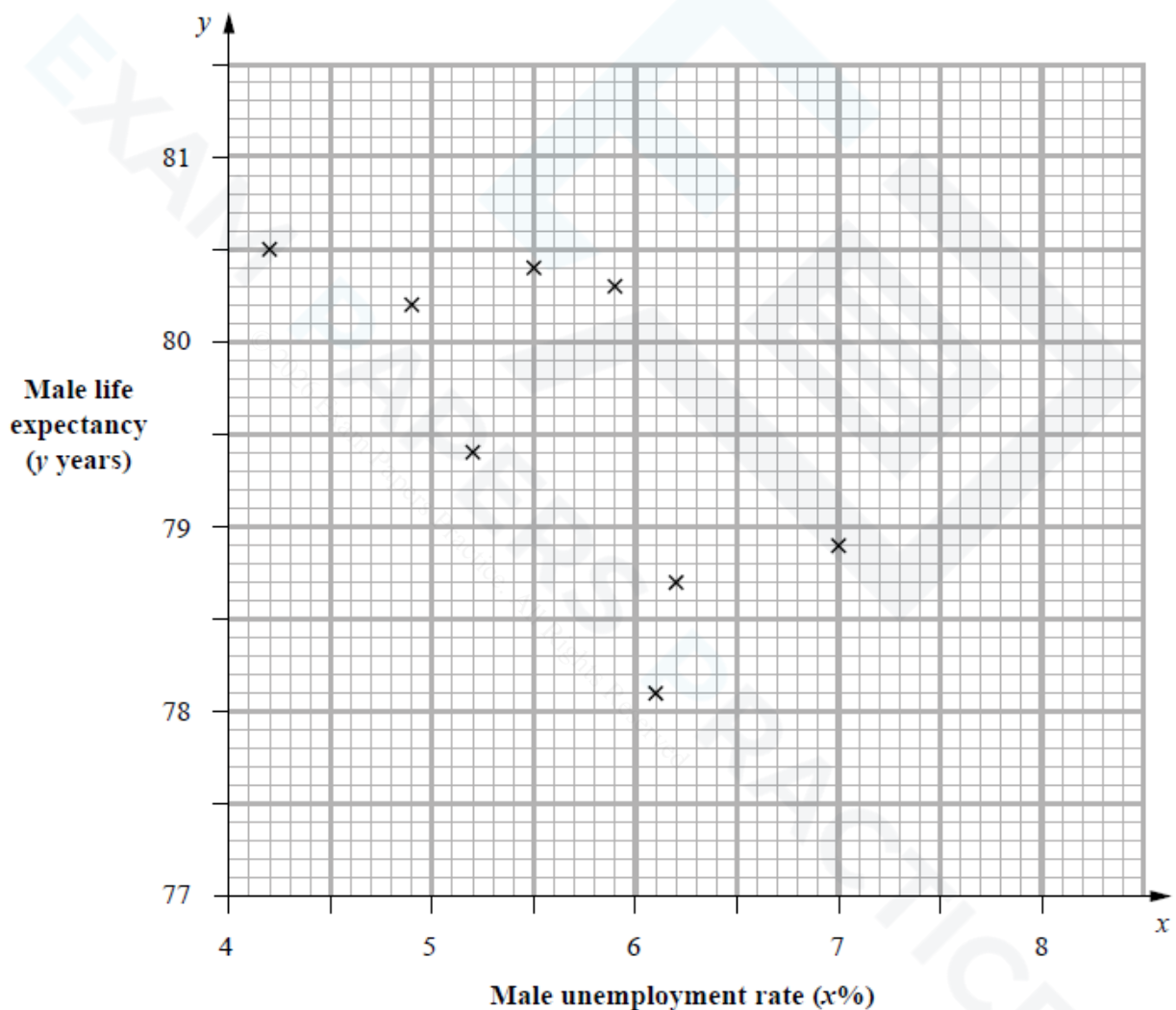
Rajesh is investigating to see if there is an association between the male unemployment rate and male life expectancy in the regions of England for 2014

His hypothesis is

"Where the male unemployment rate is high, the male life expectancy is low."

Rajesh finds the male unemployment rate ($x\%$) and the male life expectancy (y years) for each of eight of the nine regions of England for 2014

This information is shown on the scatter diagram below.



(Source: *ons.gov.uk*)

(a) Explain, giving a statistical reason, whether or not this scatter diagram supports Rajesh's hypothesis.

.....

.....

.....

(2)

The equation of the regression line for the data in the scatter diagram is $y = 83.5 - 0.7x$

(b) Draw this line on the scatter diagram.

(2)

The male unemployment rates for these eight regions have a mean of 5.6%

(c) Use this information to find the mean of the male life expectancies for these regions.

..... years

(1)

(d) Interpret the value of the gradient of this regression line.

.....

.....

.....

.....

(2)

Rajesh now finds that the male unemployment rate for the missing ninth region is 8.0%

(e) Give **two** reasons why Rajesh should not use the regression line to predict the male life expectancy in this region.

.....

.....

.....

.....

.....

(2)

(Total for question = 9 marks)

Q7.



Gary is going to investigate the amounts of time students spend watching TV.

He is going to write a plan for this investigation.

His hypothesis is

"The amount of time that boys spend watching TV is greater than the amount of time that girls spend watching TV".

Write down three other things he should include in his plan.

Explain why each of these things is appropriate.

You must refer to more than one stage of the statistical enquiry cycle.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Total for question = 6 marks)

Q8.

Anne is investigating the effect of caffeine on reaction times.