

Please check the examination details below before entering your candidate information

Candidate surname

Other names

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

Time 1 hour 30 minutes

**Paper
reference**

1ST0/2H

Statistics
PAPER 2
Higher Tier

You must have:

Ruler graduated in centimetres and millimetres, protractor,
pair of compasses, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided
– *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.



Information

- The total mark for this paper is 80
- 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.
- Good luck with your examination.

Turn over ►

P65516A

©2021 Pearson Education Ltd.

1/1/1/1/1/2 h stat-nov-21



552 Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$\text{Rates of change (e.g. Crude birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}})$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1 Ben is investigating how far people travel to work each day. He asks the people in his office how far, in miles, they travel to work each day.

The grouped frequency table gives information about their answers.

Distance (d miles)	Frequency
$0 < d \leq 5$	7
$5 < d \leq 10$	10
$10 < d \leq 15$	9
$15 < d \leq 20$	6
$20 < d \leq 25$	3

- (a) Calculate an estimate of the mean distance.
Give your answer correct to one decimal place.

..... miles
(3)

Ben finds a newspaper report that says that the average distance travelled to work each day in England and Wales in 2001 was 8.3 miles.

(Source: www.telegraph.co.uk)

From this Ben concludes that the distance travelled to work each day in England and Wales has increased since 2001

- (b) Give **two** reasons why Ben's conclusion may not be reliable.

.....

.....

.....

.....

.....

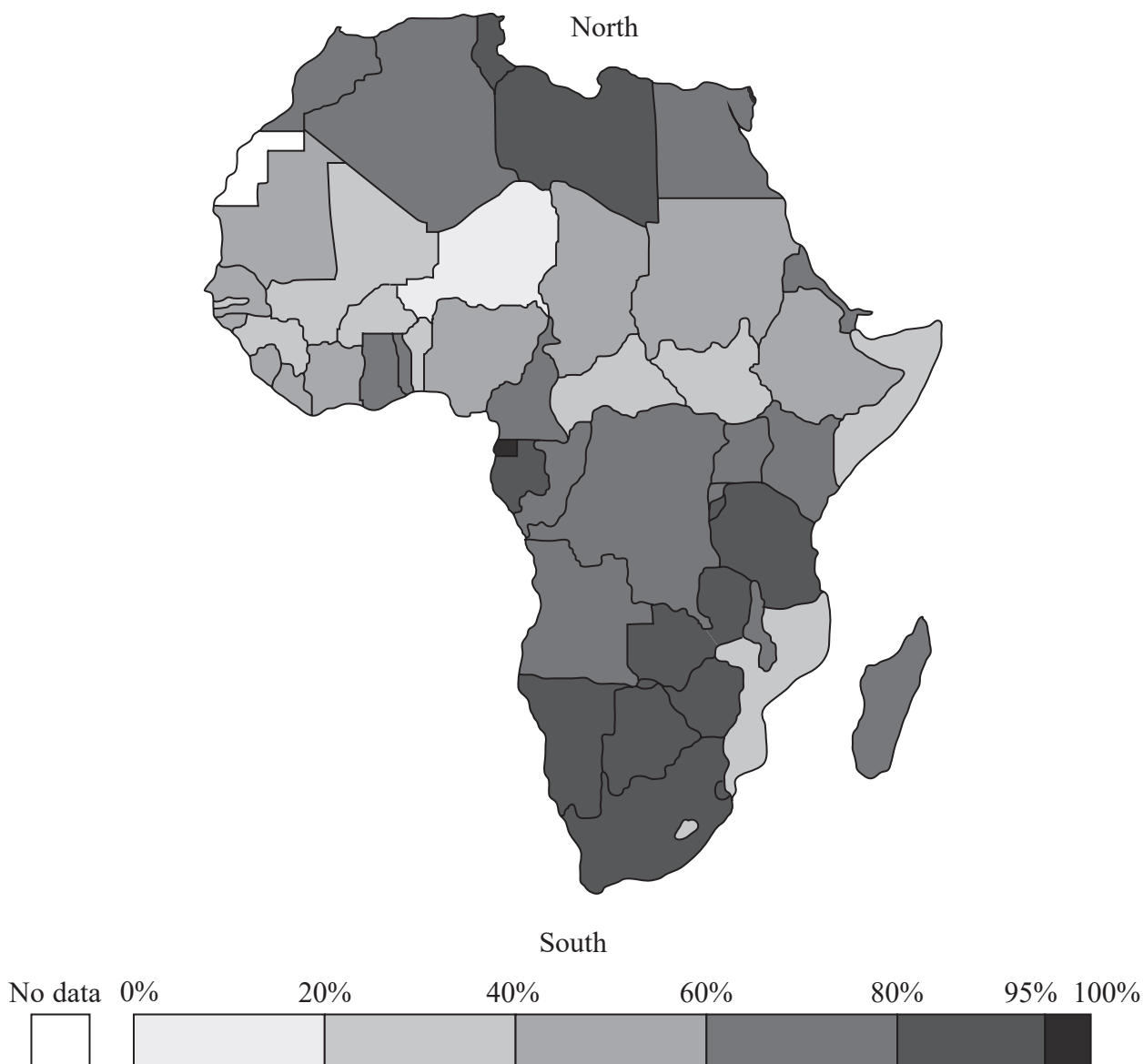
(2)

(Total for Question 1 is 5 marks)



2 Tim is writing an article for university lecturers on literacy rates in Africa.

He uses the following choropleth map to display the literacy rates in the countries of mainland Africa and Madagascar.



(a) Discuss whether or not this is a suitable diagram to use for the readers of his article.

(1)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

In his article, Tim concludes that the south of Africa has higher literacy rates than other parts of Africa.

- (b) Does the choropleth map support this conclusion?
Give statistical reasons for your answer.

.....

.....

.....

.....

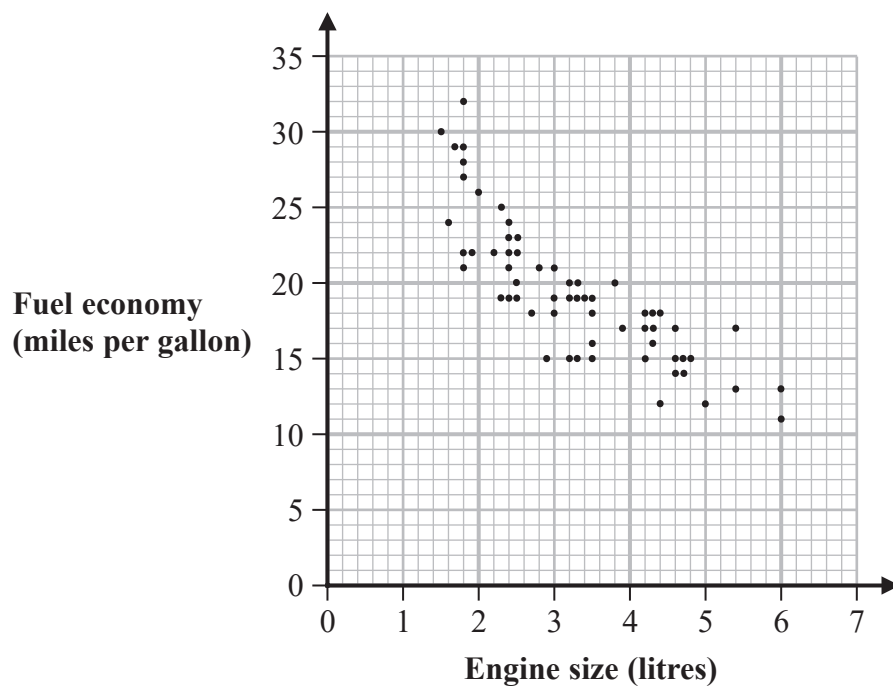
(3)

(Total for Question 2 is 4 marks)



3 Rachel collects data on the engine size (in litres) and the fuel economy during city driving (in miles per gallon) for each car in a sample of 61 cars of various types.

The scatter diagram was drawn for this information by statistical software.



(Source: *dasl.datadescription.com*)

(a) Describe and interpret the type of correlation shown by the scatter diagram.

.....

.....

.....

(2)



The statistical software worked out that the mean engine size for the 61 cars in the sample is 3.3 litres, correct to one decimal place.

The statistical software worked out that the total of the fuel economies for the 61 cars in the sample is 1183

(b) On the scatter diagram,

(i) plot with a cross (×) the double mean point (\bar{x} , \bar{y}) of the data,

(2)

(ii) draw a line of best fit through the double mean point.

(1)

Rachel wants to predict the fuel economy of a car with an engine size of 7 litres.

(c) Explain why it is not appropriate to use the line of best fit on this scatter diagram to find this prediction.

(1)

(Total for Question 3 is 6 marks)



4 Anne is investigating the effect of caffeine on reaction times.

Here is what Anne has written down for her hypothesis and how she plans to collect, process and present her data.

Hypothesis

- Does caffeine reduce reaction times?

Data collection

- Pre-test the reaction time test
- Each person to complete the reaction time test, then drink some coffee, wait a while and then complete the reaction time test a second time
- Each person to record their reaction times in the spreadsheet

Processing and presenting data

- Look for outliers in the data and remove them if found
- Calculate the median and quartiles
- Draw a box plot for the reaction times after coffee has been drunk

Discuss whether Anne’s hypothesis and her plans for collecting, processing and presenting data are appropriate.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



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)



- 5 Birgitta works in a wildlife reserve. She has been researching the size of tiger feet in order to help in the identification of animal tracks.

The table gives information about the front pad width of adult tigers.

	Front pad width	
	Mean (cm)	Standard deviation (cm)
Male tigers	11.4	0.6
Female tigers	9.2	0.4

(Source: *keywordsuggest.org*)

- (a) Compare in context the distribution of the front pad widths of adult male tigers with the distribution of the front pad widths of adult female tigers.

(2)

- (b) Explain when it would be more appropriate to use the median rather than the mean in describing a distribution.

(1)

(Total for Question 5 is 3 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



6 Emily and Jess work in the head office of a chain of restaurants.

The menu in all of the restaurants has recently been changed.
Emily and Jess have been asked to find out the effect of these changes.

They are going to start by finding out what the kitchen staff think of the new menu.

Emily thinks that they should use a cluster sample of kitchen staff in the restaurants.
The restaurants would be the clusters.

Jess thinks that they should use judgement sampling.

The head chef of the company would be asked to select suitably experienced kitchen staff for the judgement sample.

(a) Discuss whether these two sampling methods are suitable for Emily and Jess to use.

(3)

A website gives diners the opportunity to rate their dining experience at restaurants.
Here is one of the questions that diners are asked to answer.

Poor	1	2	3	4	5	Excellent
<p>Mark with a cross (X) on the scale, one of the numbers 1 (poor), 2, 3, 4, 5 (excellent) to show how you rate your meal at the restaurant.</p>						

Emily recorded the mean ratings, correct to one decimal place, for meals at each of the 28 restaurants in the chain.

The stem and leaf diagram gives information about Emily's recorded mean ratings.

1	6 7 7 9
2	3 4 5 5 7 9
3	1 1 3 4 6 6 7 8 8
4	2 3 3 5 6 6 6 7 7

Key:

3 | 4 represents 3.4



Emily picks at random one of the restaurants.

- (b) Find the probability that the mean rating recorded by Emily for meals at this restaurant is less than 2.5

(1)

Jess separates the ratings into those that were made for meals from the previous menu and those that were made for meals from the new menu.

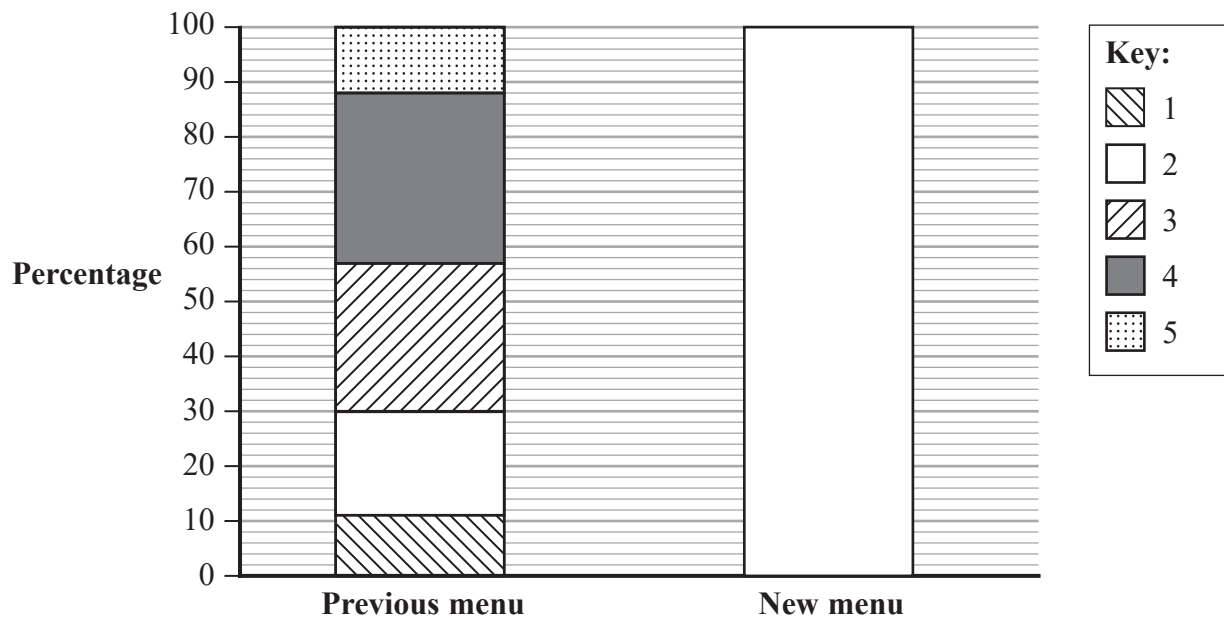
Jess uses her data to draw a percentage composite bar chart in order to compare the ratings for the previous menu and the ratings for the new menu.

The table below gives the percentages for each rating for the new menu.

Rating	1	2	3	4	5
Percentage	16	22	12	28	22

- (c) Use the information in the table to complete the percentage composite bar chart for the new menu.

(2)



- (d) Use information from the composite percentage bar charts to describe what conclusions can be made about the effects of the changes to the menu.

(2)

(Total for Question 6 is 8 marks)



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

(a) Suggest a hypothesis Matthew could use.

(1)

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 is 7 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

8 David is writing a project about iguanas.

He finds data on the internet posted by a biologist investigating iguanas on the Galapagos Islands.

The biologist captured and tagged a sample of iguanas in La Loberia.
The tagged iguanas were then released back from where they were captured.

After a few days the biologist captured at random a second sample of iguanas in La Loberia and recorded the number of tagged iguanas in this sample.

Here is the biologist's data.

Number of iguanas in first sample	Number of iguanas in second sample	Number of tagged iguanas in second sample
114	243	69

(Source: www.discoveringgalapagos.org.uk)

(a) Work out an estimate of the total number of iguanas in La Loberia.

.....
(2)

(b) Discuss the reliability of **using the biologist's data** to work out an estimate of the total number of iguanas in La Loberia.

.....
(2)

It is suggested that some of the tags might have fallen off the iguanas between the two samples.

(c) If this had happened, describe the effect it would have on the estimate worked out in part (a).

.....
(1)

(Total for Question 8 is 5 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

9 A construction company employs three types of employee: builders, carpenters and electricians.

The table gives the number of each type of employee.

Type of employee	Number
Builders	48
Carpenters	26
Electricians	22

The owner of the construction company wants to find out what the employees feel about their jobs.

She plans to take a sample of 30 employees, stratified by type of employee.

(a) Calculate the number of carpenters that should be in her sample.

.....
(2)

The owner wants to know whether the employees in her sample believe that they work hard consistently.

She knows that she will have to ask a sensitive question so she plans to use a random response question in order to find out this information.

(b) Design a random response question that the owner could use on a questionnaire in order to collect this information.

(4)

(Total for Question 9 is 6 marks)



- 10 The table gives the standardised scores for Brinn Bevan in three of the gymnastic apparatus used in the men's all-around final in the 2018 Artistic Gymnastics World Championships.

Apparatus	Floor exercise	Pommel horse	Rings
Standardised score	0.763	-0.938	-0.0169

(Source: *worldgymdoha18.com*)

- (a) Use the standardised scores to compare Brinn Bevan's performances **relative to the other competitors** in the floor exercise, in the pommel horse and in the rings. Give reasons for your comparisons.

(2)

The raw score for each apparatus is also known.
Each apparatus had the same maximum raw score.

- (b) Explain why it is appropriate to use standardised scores rather than raw scores to compare performances.

(1)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

The table below shows Brinn Bevan's standardised score for the parallel bars.

It also shows the mean and the standard deviation of the number of points awarded in the parallel bars to all the gymnasts who completed the men's all-around final.

	Standardised score	Mean	Standard deviation
Parallel bars	0.247	14.389	0.854

(Source: *worldgymdoha18.com*)

- (c) Work out the number of points awarded to Brinn Bevan for the parallel bars.
Give your answer correct to one decimal place.

.....
(2)

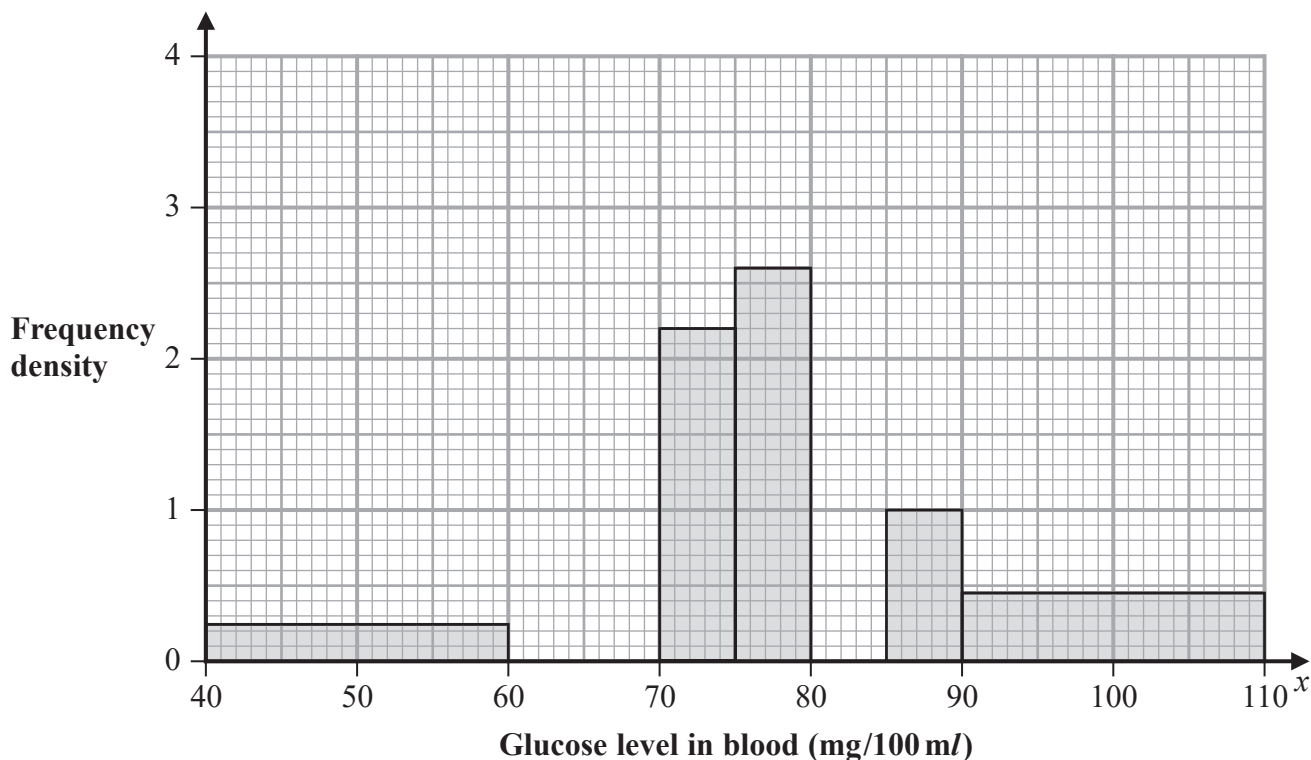
(Total for Question 10 is 5 marks)



11 The glucose level in blood is measured in milligrams per 100 millilitres (mg/100ml).

A doctor is investigating the glucose levels in blood of some adult females at the end of 12 hours of not eating or drinking.

The incomplete histogram and the incomplete grouped frequency table give information about the glucose levels in the blood of these adult females.



Glucose level in blood x (mg/100 ml)	Frequency
$40 < x \leq 60$	5
$60 < x \leq 70$	11
$70 < x \leq 75$	11
$75 < x \leq 80$	13
$80 < x \leq 85$	16
$85 < x \leq 90$	
$90 < x \leq 110$	

(Source: college.cengage.com)



(a) Use the information in the histogram to complete the table. (2)

(b) Use the information in the table to complete the histogram. (2)

The doctor used statistical software to produce the following summary statistics for his data.

$$n = 70 \quad \sum x = 5459 \quad \sum x^2 = 436\,137$$

(c) By calculating limits for outliers using the mean and the standard deviation, explain whether or not there could be any outliers in the doctor's data.

(5)

A second doctor is investigating the glucose levels in the blood of some adult males at the end of 12 hours of not eating or drinking.

The following summary statistics were calculated.

Mean	Median	Standard deviation
75.6	79	11.8

(d) Calculate the skew for the distribution of the adult male glucose levels in blood and interpret this value in context.

Give your answer correct to 2 decimal places.

(3)

(Total for Question 11 is 12 marks)



12 A company contracts two manufacturers, **A** and **B**, to make an item that the company sells.

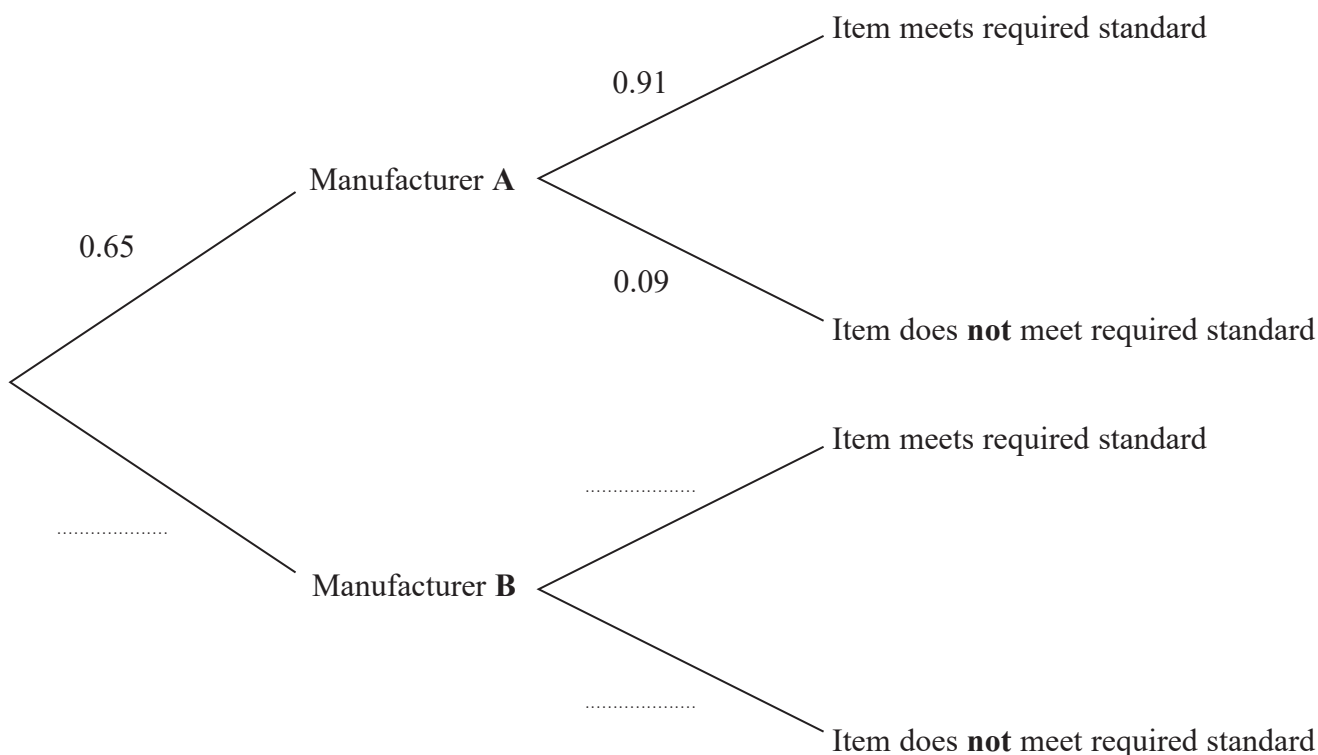
Manufacturer **A** makes 65% of the items for the company to sell.

It is found that 9% of the items made by manufacturer **A** do not meet the company's required standard.

It is found that 12% of the items made by manufacturer **B** do not meet the company's required standard.

An item is to be picked at random.

(a) Complete the probability tree diagram for this item.



(1)

One of the items that the company is to sell is picked at random.

M is the event that the item was made by manufacturer **A**.

N is the event that the item does **not** meet the required standard.

(b) Find $P(M|N)$.

(2)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

The company is planning to sell multipacks containing 6 items made by manufacturer A.

The financial director of the company plans to use the binomial distribution to model the number of items in the multipack not meeting the required standard.

- (c) Write down one condition that needs to be assumed so that the binomial distribution is a suitable model to use.

(1)

A multipack made by manufacturer A is to be selected at random.

- (d) Work out the probability that this multipack will have 2 or more items that do not meet the required standard.
Give your answer correct to 3 decimal places.

(3)

(Total for Question 12 is 7 marks)



- 13 A team of researchers investigated the time intervals between calls of tree frogs.

They investigated two species of tree frog: the Italian tree frog and the European tree frog.

They found that there was a correlation between air temperature, $x^{\circ}\text{C}$, and amount of time between calls, y milliseconds, for each of the two species of tree frog.

The table gives the equation of the regression line for the data for each of the two species of tree frog.

Species	Explanatory variable (x)	Response variable (y)	Equation of regression line
Italian tree frog	Temperature ($^{\circ}\text{C}$)	Time between calls (milliseconds)	$y = 182.3 - 5.87x$
European tree frog	Temperature ($^{\circ}\text{C}$)	Time between calls (milliseconds)	$y = 237.2 - 7.90x$

(Source: www.zobodat.at)

When $x = 27$, both equations give the same value of y , correct to the nearest whole number.

- (a) Explain what this means in context.

(1)

The equation of the regression line gives information about the relationship between temperature and time between calls for each of the two species of tree frog.

- (b) Using this information, compare the relationship for the Italian tree frog with the relationship for the European tree frog.

(5)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

A second group of researchers investigated the relationship between temperature and the time between calls for a third species of tree frog. They also obtained the equation of the regression line for this species of tree frog.

The results of the different teams of researchers are to be compared.

(c) Give one potential limitation of doing this.

.....

.....

.....

(1)

(Total for Question 13 is 7 marks)

TOTAL FOR PAPER IS 80 MARKS



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

