Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
TOTAL	

General Certificate of Secondary Education
HIgher Tier
Specimen Paper

Statistics

XXXX/H

Date: Time

H

For this paper you must have:
- a calculator
- mathematical instruments

Time allowed
- 2 Hours.

Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Answers written in margins will not be marked.
- Use a calculator where appropriate.
- Do all rough work in this book.

Information
- The maximum mark for this paper is 100.
- The marks for questions are shown in brackets.
- You may ask for more answer paper and graph paper. These must be tagged securely to this answer book.

Advice
- In all calculations, show clearly how you work out your answer.

Answer **all** questions in the spaces provided

1 A National Census form contains questions to be answered by each household and each person. Below are lists giving the subjects of some of the questions on a National Census form.

Household

A1 Address
A2 Ownership
A3 Number of rooms
A4 Number of cars
A5 Washing, cooking and toilet facilities

Each person

B1 Name
B2 Date of birth
B3 Sex
B4 Usual address
B5 Relation to head of household
B6 Whether married
B7 Whether the person has a job
B8 Work or full time education
B9 Place of Birth
B10 Parents' place of birth
B11 Address 1 year ago
B12 Address 5 years ago
B13 Academic qualifications
B14 Professional qualifications
B15 Employer
B16 Type of work
B17 Employee or self-employed
B18 Whether an apprentice
B19 Hours of work
B20 Place of work
B21 Method of travel to work
B22 Job last year
B23 Birth of children
B24 Marriage date

For each of the following pieces of information put down one or more items from the table that could help to supply these answers.

(For example B21 would help us to find the number of people who drive to work.)

(a) The average number of rooms per household

Answer .....

(1 mark)

(b) The number of people who moved in the last year.

Answer .....

(2 marks)

(c) The number of married women, with children, working in the car industry.

Answer.....

(3 marks)

(d) Peter wishes to conduct a survey in his constituency to estimate the proportion of Labour voters in the area.

The constituency is split into five wards, each ward has six polling districts.

All eligible voters are registered with their local polling district.

Peter needs to select his sample of voters from one polling district only.

(i) Describe how Peter could select the sample of voters using multi stage sampling methods.

.....

.....

.....

(3 marks)

(ii) Peter will contact his sample of voters on-line

Give one advantage and one disadvantage in using this method of data collection.

Advantage .....

.....

Disadvantage .....

.....

(2 marks)

(e) Results from a pilot survey showed 58% of voters supporting Labour.

Use this value to estimate the proportion of Labour voters in the constituency.

.....

Answer .....

(1 mark)

12

2 The table shows the percentage by age for each ethnic group of the UK population 2007 - 2008

Ethnic Group \ Age	Under 16	16 – 34	35 – 64	65 and over
White	19	25	40	16
Mixed	55	27	16	2
Indian	22	34	38	6
Pakistani	35	36	25	4
Bangladeshi	38	38	21	3
Other Asian	22	36	38	4
Black Caribbean	24	25	42	9
Black African	33	35	30	2
Other Black	35	34	26	5
Chinese	20	40	35	5
Other	20	37	39	4

Source: *Adapted from Office for National Statistics, Summer 2008*

(a) Which ethnic group had the largest percentage of its population under 16 years of age?

Answer .....

(1 mark)

(b) What was the difference between the percentages of Chinese ethnic group and Black African ethnic group aged 35 - 64 years?

.....

Answer .....

(2 marks)

Barcode

LEAVE MARGIN BLANK

(c) Give one similarity and one difference between the age profiles of the White ethnic group and the Indian ethnic group.

Similarity .....

.....

Difference .....

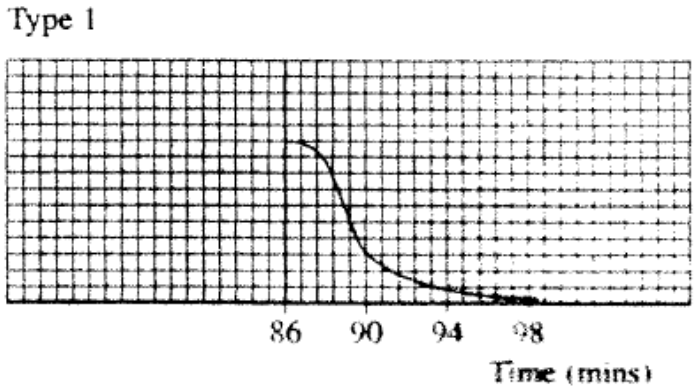
.....

(2 marks)

5

3 A company produces equal numbers of two types of engine component. The time taken to produce Type 1 components form a normal distribution.

(a) On the grid below complete the diagram for the frequency distribution.



(2 marks)

(b) The times for the Type 2 components are also normally distributed with mean 92 minutes and standard deviation 2 minutes. On the same grid draw a diagram to represent this frequency distribution.

(3 marks)

(c) For Type 1, what proportion of the engine components will take 86 minutes or less to complete?

.....

(1 mark)

(d) For Type 2, what proportion of the engine components will take more than 100 minutes to complete?

.....

(1 mark)

7

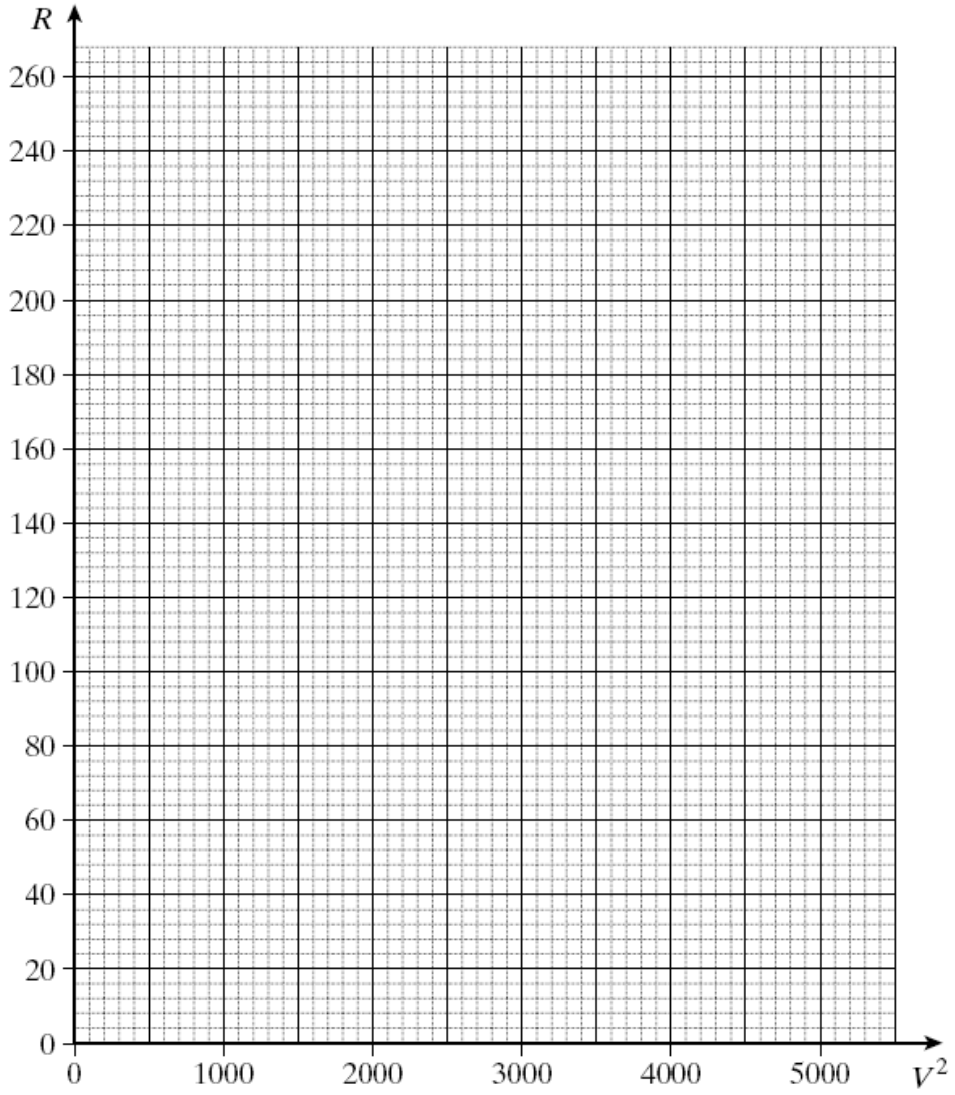
Barcode

LEAVE MARGIN BLANK

4 The speed of a car, V km/h, and the total resistance to its motion, R newtons, are given in the table.

V	10	20	30	40	50	60	70
R	35	45	65	90	130	170	225
V^2							

(a) Plot a graph or R against V^2.



(3 marks)

(b) The mean of V^2 is 2000 and the mean of R is 109.
Draw a line of best fit on your graph.

(2 marks)

(c) Write down the intercept with the R axis.

Answer .....

(1 mark)

(d) Calculate the gradient of your line of best fit.

.....

.....

Answer .....

(3 marks)

(e) Write down a formula for R in terms of V^2.

.....

.....

Answer .....

(1 mark)

(f) Use your formula to estimate the resistance to motion when the speed of the car is 90 km/h

.....

.....

Answer ..... newtons

(2 marks)

(g) Is this estimate reliable?
Give a reason for your answer.

.....

.....

(1 mark)

13

LEAVE MARGIN BLANK

5 An expert from the local antiques club agreed to challenge a number of contestants to correctly rank eight items of Victorian furniture according to their value.

John agreed to take part and his **rankings** along with those of the expert were as follows

Exhibit	A	B	C	D	E	F	G	H
Expert	1	3	6	7	8	2	5	4
John	8	5	2.5	4	1	6	2.5	7

(a) Calculate the value of Spearman's rank correlation coefficient for the two sets of data.

.....

.....

.....

.....

.....

Answer.....

(4 marks)

(b) Interpret, in context, your answer to part (a).

.....

.....

(1 mark)

(c) A further eight contestants entered the competition.
The values of the correlation coefficients were

0.35 -0.43 0.71 0.05 -0.36 -0.02 0.92 -0.81

Which two of these values show that there is almost no correlation between the individual rankings of that contestant and those of the expert?

Answer.....

(2 marks)

(d) Explain why the Product Moment Correlation coefficient would not be an appropriate measure of correlation in this case

.....

.....

(1 mark)

8

Barcode

LEAVE MARGIN BLANK

6 A firm produces tins of baked bean. For quality control purposes a sample of five tins of baked beans is taken every hour and the mass of each tin is measured.

The mean mass and range of masses of each sample is calculated and plotted on separate graphs.

The graphs below show the mean mass and range of masses of the first seven samples. The eighth sample has tins of baked beans of the following masses:

1.072 kg 0.998 kg 1.024 kg 1.037 kg 1.046 kg

The mean mass of this sample is 1.0354 kg.

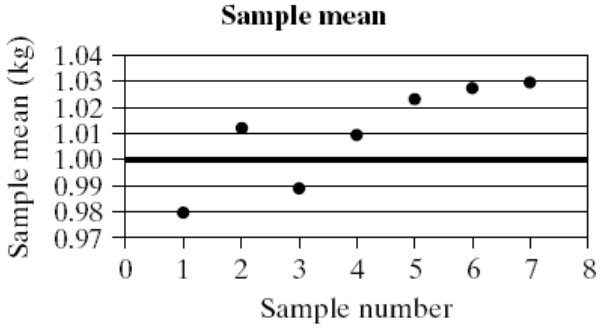
(a) Calculate the range of masses of this sample.

Answer..... kg

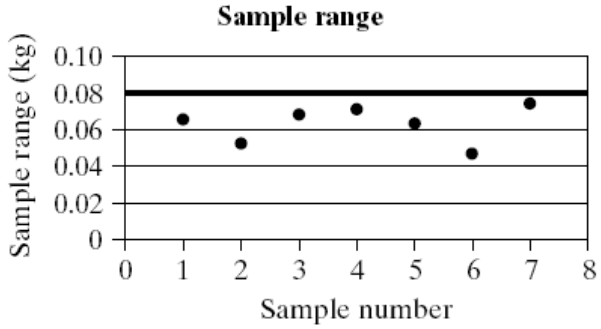
(1 mark)

(b) Plot the values of the eighth sample on the appropriate graphs.

The target mean is 1 kg.



The acceptable range is 0.08 kg.



(2 marks)

(c) Make **one** comment on each graph in relation to the production process.

Mean.....

.....

Range .....

.....

(2 marks)

5

7 Staywell medical practice has 6 doctors.
Each doctor is equally likely to be on call.

(a) Describe how you could use a dice to simulate the selection of a doctor.

.....

.....

(1 mark)

(b) Get Better Medical practice B collected data throughout 2008
The following table shows the average number of patients, by age group, attending Get Better medical practice B per day.

Age	Patients
Under 16	38
16 to 59	35
60 and over	27
Total	100

Explain how you could use random numbers to simulate the ages of the next two people attending the medical practice.

.....

.....

.....

.....

(3 marks)

4

Barcode

LEAVE MARGIN BLANK

8 The diagram shows a factory layout divided into four different work areas.

The numbers of male and female staff in each area are also given.

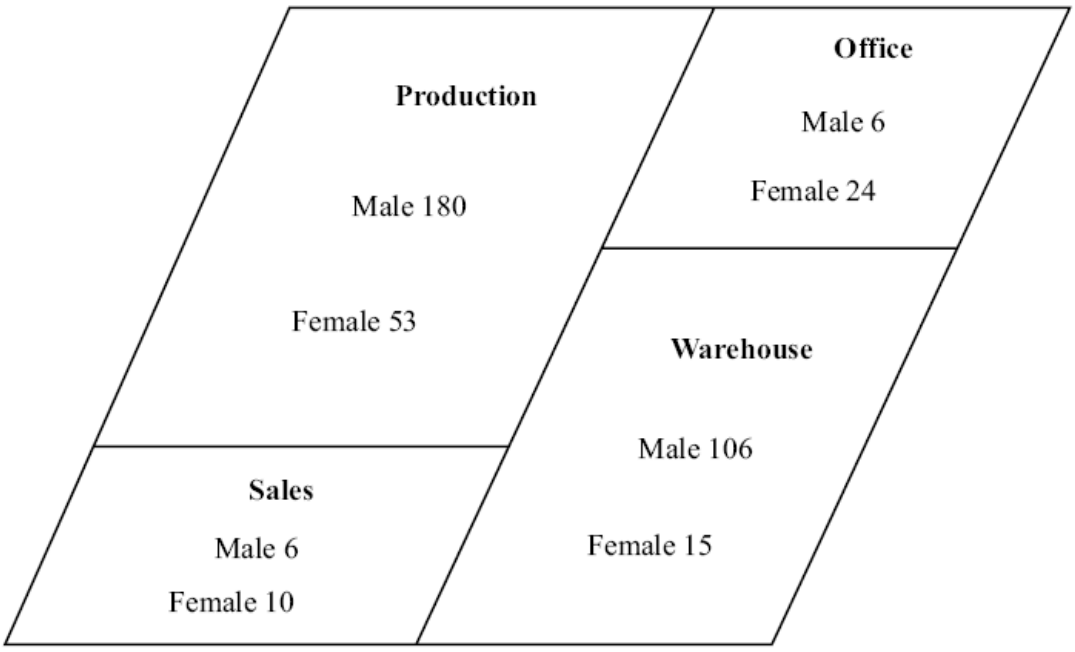


Diagram not to scale

The factory owner wishes to undertake a survey to find the reaction of the staff to the introduction of a new bonus scheme.

He decides to take a systematic sample of 20 male production staff.

(a) Explain how this sample could be collected.

.....

.....

.....

.....

(3 marks)

(b) Give **two** reasons why this sample would be unrepresentative of the whole staff.

Reason 1 .....

.....

Reason 2 .....

.....

(2 marks)

(c) As an alternative the owner is advised to take a sample, stratified by sex and work area, of 50 of the 400 staff.

(i) Calculate the number of sales staff to be included in the sample.

.....

.....

Answer .....

(2 marks)

(ii) Calculate the number of female office staff to be included in the sample.

.....

.....

Answer .....

(2 marks)

LEAVE MARGIN BLANK

(d) Part of the survey will involve interviewing staff to find their views on plans to change the number of hours worked each week in the factory.

One of the questions to be asked will be

> *What is your opinion of the proposal to increase the number of hours worked each week to 39?*

Describe two types of scale that could be used to measure the opinions in this case.

Type 1 .....

.....

.....

.....

Type 2 .....

.....

.....

.....

(4 marks)

13

9 Ten pupils took an exam.

$\sum x = 315$ (where x is the number of marks recorded for each pupil)

$\sum x^2 = 10829$

$\sum \left(x - \bar{x}\right)^2 = 906.5$

(a) Calculate the mean and standard deviation for these ten exam marks.

.....

.....

.....

.....

.....

.....

.....

.....

Mean .....

Standard deviation .....

(4 marks)

Barcode

LEAVE MARGIN BLANK

(b) One of the marks was 32 but was incorrectly recorded as 23.

The corrected mean and standard deviation were 32.4 and 9.09

Explain

(i) why the mean has increased,

.....

.....

(1 mark)

(ii) why the standard deviation has decreased.

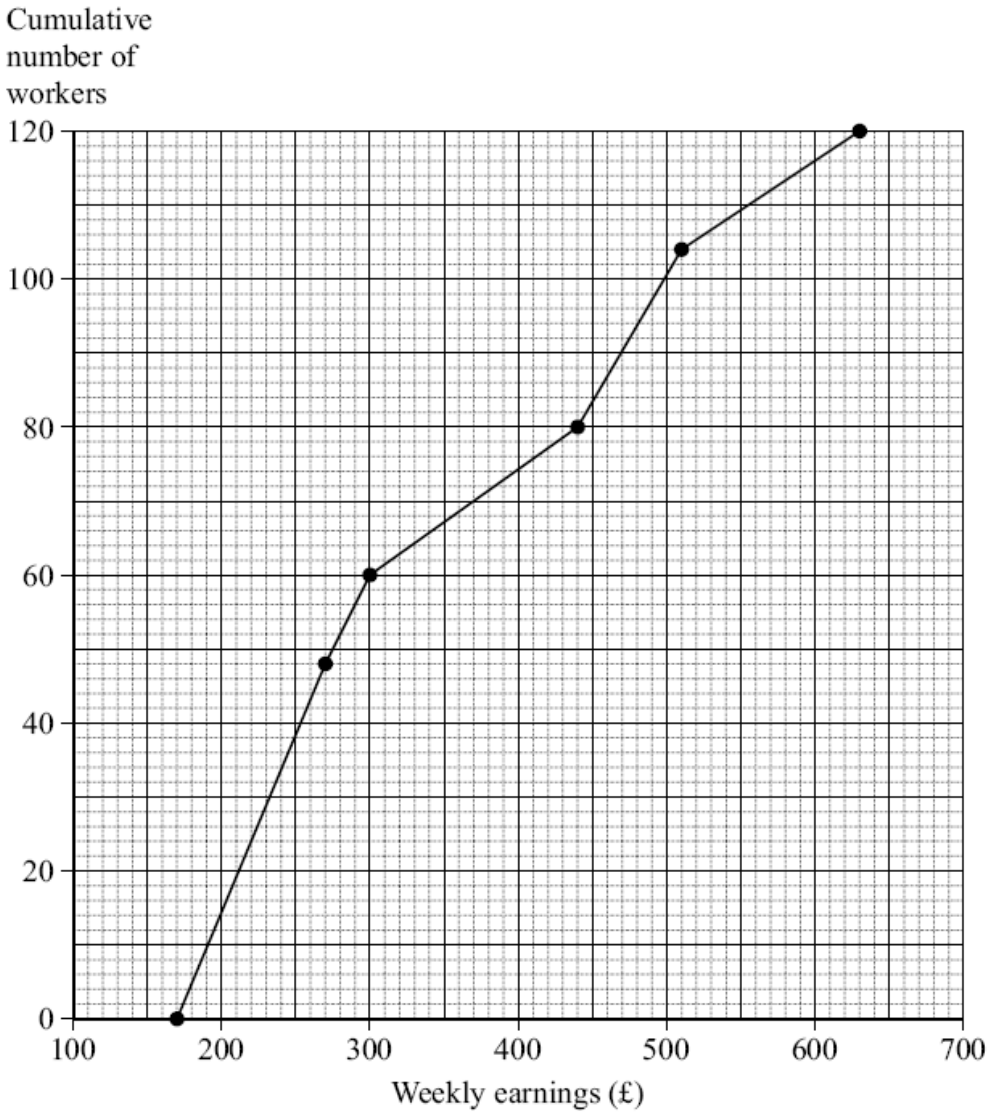
.....

.....

(1 mark)

6

10 The cumulative frequency polygon shows the distribution of weekly earnings of a sample of 120 male manual workers in the ceramics industry.



Barcode

LEAVE MARGIN BLANK

(a) Use the graph to estimate

(i) the median

Answer £.....

(1 mark)

(ii) the interquartile range

.....

.....

Answer £.....

(2 marks)

(iii) the percentage of workers earning under £320 per week

.....

.....

.....

Answer.....%

(3 marks)

(iv) the 9th decile.

.....

.....

.....

Answer £.....

(2 marks)

(b) The following information was found from a sample of 120 female manual workers in the ceramics industry.

The median of the weekly earnings was £230.

25% of the sample had weekly earnings more than £280.

The interquartile range was £100.

No one earned less than £120 per week or more than £420 per week.

Six workers earned more than £390 per week.

(i) Use your answers to part (a) and the information on female earnings to make **two** statements that support the following hypothesis:

'Female workers in the ceramics industry have lower **and** less variable weekly earnings than male workers in the ceramics industry.'

Statement 1 .....

.....

Statement 2 .....

.....

(2 marks)

(ii) Describe another source of data that could be used to explore this hypothesis.

.....

.....

(1 mark)

LEAVE MARGIN BLANK

(c) One method of calculating a measure of skewness is:

$$\frac{\text{Upper Quartile + Lower Quartile} - 2\text{ (median)}}{\text{Upper Quartile} - \text{Lower Quartile}}$$

(i) Use this method to calculate separate measures of skewness for the weekly earnings of male **and** female workers.

.....

.....

.....

.....

.....

.....

.....

Answer Male workers .....

Female workers .....

(3 marks)

(ii) Use your answers to part (c)(i) to describe the shape of each distribution.

Answer Male workers .....

Female workers .....

(2 marks)

16

11 Records for a local library show for each book whether it is in the fiction, non fiction or classics category and whether it is a hard back or soft back version.

When the library closed on Wednesday last week 2700 books were out on loan.

Of the books on loan 72% were in the fiction category.

Of the 620 hard back books on loan 55% were in the non fiction category and 25% in the classics category.

In total 176 classics books were on loan.

(a) Complete the table, entering the number of books on loan in each case.

Version / Category	Hard back	Soft back	Totals
Fiction			
Non Fiction			
Classics			176
Totals	620		2700

(4 marks)

(b) A library record for a book on loan is chosen at random.

Use the table to calculate the probability that the book is

(i) non fiction and a soft back version

.....

Answer .....

(1 mark)

(ii) either a non fiction or a hard back version or both

.....

Answer .....

(2 marks)

(iii) fiction, given that it is a soft back version.

.....

Answer .....

(2 marks)

Barcode

LEAVE MARGIN BLANK

(c) How many of the first 200 books taken out on loan on the following day would you expect to be hard back classics?

.....

.....

.....

Answer.....

(2 marks)

11