OCR
Oxford Cambridge and RSA

Level 3 Certificate
Quantitative Problem Solving (MEI)
H867/02 Statistical Problem Solving

Wednesday 23 May 2018 – Morning
Time allowed: 2 hours

You must have:
- the Insert (inserted)
- the Statistical Tables (ST1) (inserted)

You may use:
- a scientific or graphical calculator

First name	
Last name	
Centre number	Candidate number

INSTRUCTIONS
- The Insert will be found inside this document.
- Use black ink. You may use an HB pencil for graphs and diagrams.
- Complete the boxes above with your name, centre number and candidate number.
- Answer **all** the questions.
- Write your answer to each question in the space provided. If additional space is required, you should use the lined page(s) at the end of this booklet. The question number(s) must be clearly shown.
- Do **not** write in the barcodes.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

INFORMATION
- The total mark for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- This document consists of **20** pages.
- Final answers should be given to a degree of accuracy appropriate to the context.

Answer **all** the questions.

Section A (30 marks)

1 A house building company applies for planning permission to build houses on a flood plain just outside a small town. Elaine is a journalist with the local newspaper. She wants to write a fair and informed report about the proposed development.

She commissions 5 assistants. In order to obtain the views of 100 people, each of them is to ask a sample of 20 people, 10 male and 10 female, two questions.

- Do you support the proposed housing development?
- What is the main reason for your view?

(i) Which of the following terms describes the sample best?

Opportunity, Simple Random, Stratified, Quota, Cluster, Self-selected. **[1]**

The results for the first question are summarised in the table below.

	Male			Female		
Interviewer	**Yes**	**No**	**Don't know**	**Yes**	**No**	**Don't know**
A	6	3	1	4	4	2
B	5	5	0	3	5	2
C	4	6	0	5	2	3
D	3	5	2	3	3	4
E	5	4	1	3	4	3
Total	**23**	**23**	**4**	**18**	**18**	**14**

(ii) State two general points that Elaine can conclude from the figures in the table. **[2]**

The most common reasons given are:

For: We need more housing in this town; there's nowhere for young people to live (except with their parents).

Against: Building on a flood plain means that places further down the river are more likely to be flooded.

(iii) For each of these reasons, make one suggestion as to what further data Elaine should try to collect.

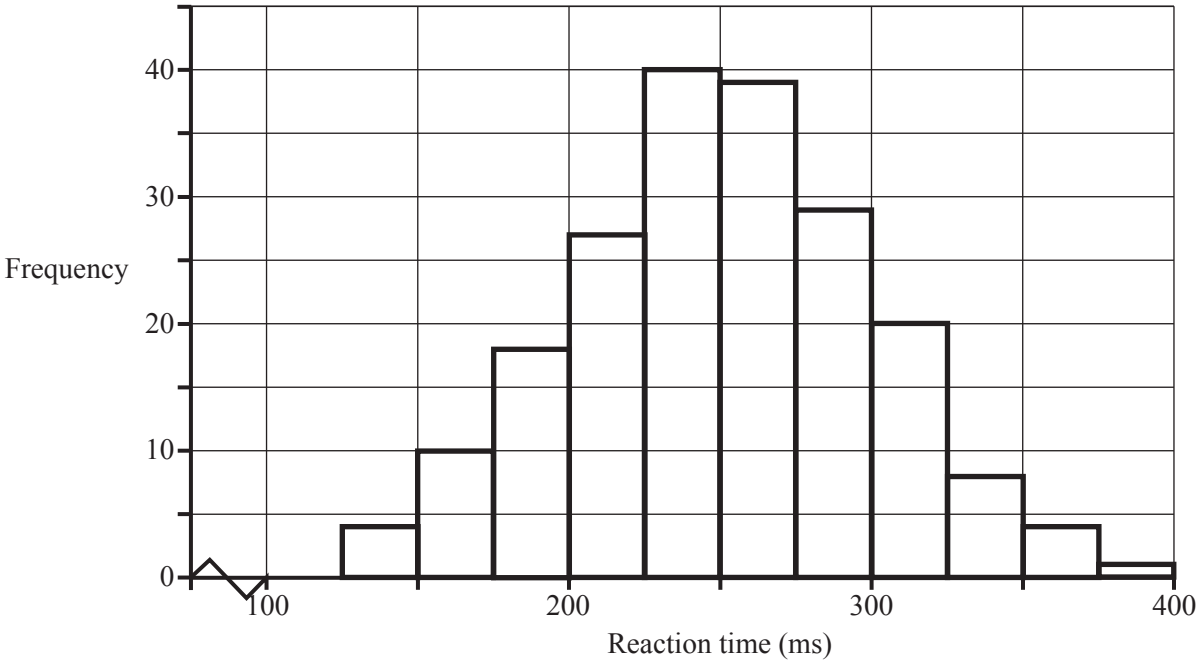
In each case say how she might obtain the data. **[4]**

1 (i)	
1 (ii)	
1 (iii)	**For**
	Against

2 Miranda is an administrator in a company that develops and hosts online tests. The company develops a game that measures reaction times. When it is nearly ready, Miranda, who has not been involved in its design, is asked to trial it.

Her first attempt gave her reaction time to be 520 milliseconds. Her next four attempts, in order, gave times of 415, 352, 242 and 268 ms.

Miranda then tries the game 200 times more. Her times for these 200 attempts are recorded and displayed as the frequency chart below. It is suggested that they can be modelled by a Normal distribution.



(i) Show that for a Normal distribution with mean 250 and standard deviation 50, the probability of an observation being between 250 and 275 is 0.1915. **[3]**

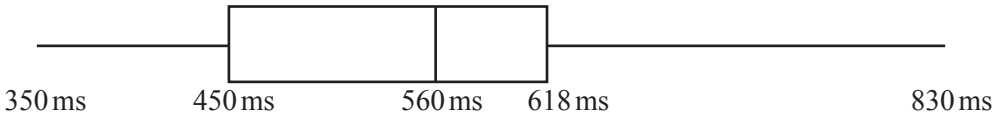
2 (i)

(ii) The equivalent figures for some other intervals are given in the table below.
Complete this table.
Give two reasons why the information in this table and the frequency chart on the previous page indicates that this Normal distribution is indeed a good model for Miranda's times. **[3]**

2 (ii)

Interval	Probability	Normal model frequency	Miranda's frequency
250 – 275	0.1915	38.30	39
275 – 300	0.1499	29.98	29
300 – 325	0.0918	18.37	20
325 – 350	0.0441	8.81	8
Over 350			5

Miranda talks about the test to two friends drinking wine at a pub. After closing time they go to her house and play the game. In total they do 40 tests; their times are shown on the box and whisker plot below.



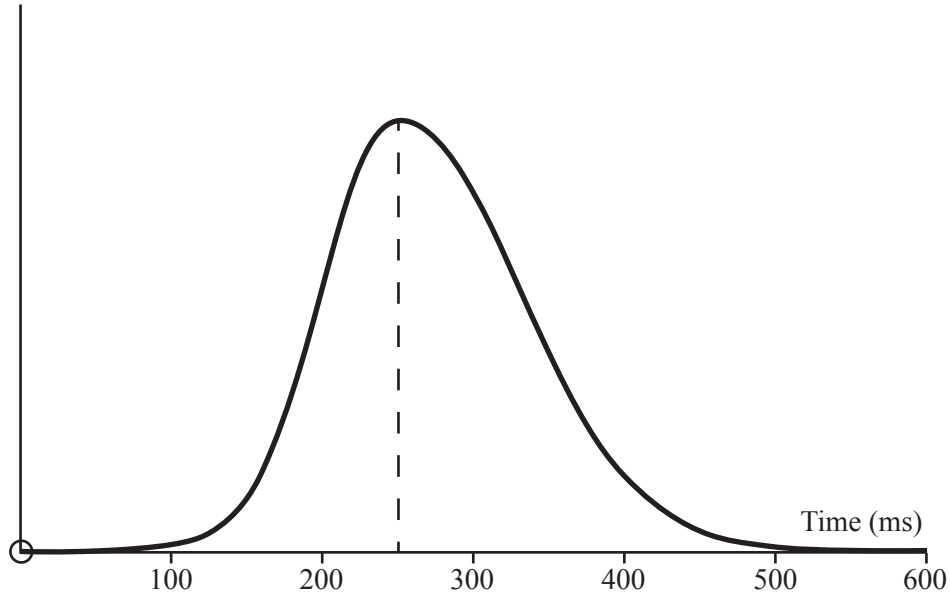
(iii) Compare these times with Miranda's times. Give a possible explanation for any difference. **[2]**

2 (iii)

(iv) Miranda's company plan to keep a record of all the times of people playing the game and to show the overall distribution in a diagram. The graph below is their prediction of what it will look like when the data come in.

(*A*) Identify its main features. **[2]**

(*B*) Give one possible explanation for the shape of the curve. **[1]**



2 (iv)(*A*)	
2 (iv)(*B*)	

3 Salim is a doctor. One of his patients gives him a bottle of home-made medicine made from wild plants using an old family recipe.

Sometimes the patient gives himself a dose of the medicine. He says his legs are a bit stiff the next day but otherwise he feels fine.

Salim obtains permission to do a pilot study on possible side effects of the medicine. He invites everyone who works at his surgery to take part and 10 healthy people volunteer.

- Before they take the medicine, they are timed running 100 m. These times, in seconds, are denoted by t_1.
- A few days later, they are given measured doses of the medicine and timed again the following day. These times are denoted by t_2.

The results are given in Table 3.1.

Volunteer	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Dose (ml)	1.00	2.00	0.00	1.50	0.50	1.75	0.25	0.75	1.25	2.25
t_1 (s)	12.3	11.6	12.8	13.4	15.1	11.2	12.3	17.5	16.3	14.4
t_2 (s)	12.0	13.0	12.8	14.1	14.5	12.4	12.2	17.1	16.9	14.9

Table 3.1

(i) Which of the following terms describes the sample best?

Opportunity, Simple Random, Stratified, Quota, Cluster, Self-selected. **[1]**

3 (i)	

Salim wants to investigate whether there is any relationship between the dose and the change in times between the first and second runs. He uses the figures in Table 3.1 to carry out a test based on Spearman's Rank Correlation Coefficient at the 5% significance level.

(ii) State the null and alternative hypotheses for this test.

Complete Table 3.2 and carry out the test. State the result. **[8]**

3 (ii)

Volunteer	Dose	Dose rank	$t_2 - t_1$	$(t_2 - t_1)$ rank	d	d^2
A	1.00	5	–0.3			
B	2.00		1.4	10		
C	0.00	1	0.0			
D	1.50		0.7			
E	0.50		–0.6	1		
F	1.75		1.2			
G	0.25		–0.1			
H	0.75		–0.4			
I	1.25		0.6			
J	2.25	10		6		16
					Σ	

Table 3.2

A week later the volunteers are timed on a third 100 m run. These times are denoted by t_3.
Their three times are given in Table 3.3.

Volunteer	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
t_1 **(s)**	12.3	11.6	12.8	13.4	15.1	11.2	12.3	17.5	16.3	14.4
t_2 **(s)**	12.0	13.0	12.8	14.1	14.5	12.4	12.2	17.1	16.9	14.9
t_3 **(s)**	12.3	11.5	12.9	13.6	15.0	11.2	12.3	17.7	16.2	14.4

Table 3.3

(iii) Salim has to write a short report for his practice manager, commenting on the pilot study.

Give three points that the report might contain. **[3]**

3 (iii)

Section B (30 marks)

The questions in this section are based on the pre-release data. A hard copy of this is provided with this examination paper.

4 **(i)** The GDP per capita is given for 228 countries in the pre-release data.

Find the median value and show that Ecuador, Macedonia and Azerbaijan are the countries with GDP per capita closest to it. **[3]**

(ii) Find the figure half way between the lowest and highest values of the GDP per capita and compare it with the median value. What does this tell you about the distribution of wealth in the world? **[3]**

4 (i)	
4 (ii)	

5 **(i)** Using relevant figures from the pre-release data set, estimate the number of babies born in a year in Argentina. **[3]**

A different measure of birth rate is considered. It is the number of babies born in a year per female aged between 15 and 54 (inclusive). In Argentina there are 11 692 613 females in this age range.

(ii) Calculate the new measure. **[2]**

(iii) The range 15 to 54 covers 40 years. Multiply your answer to part **(ii)** by 40.
What information does this give you? **[2]**

5 (i)

5 (ii)	
5 (iii)	

6 Dipali wants to know if it is more healthy to live on an island or in a mainland country. To investigate this she starts by selecting 20 islands at random (from the pre-release data) and 30 mainland countries.

She classifies these countries according to whether their life expectancy is Low, Medium or High, using her own scale.

The results are shown in Table 6.1.

f_o	**Low**	**Medium**	**High**	**Total**
Islands	1	9	10	**20**
Mainland countries	12	11	7	**30**
Total	**13**	**20**	**17**	**50**

Table 6.1

Dipali uses the data in Table 6.1 to carry out a χ^2 test at the 5% significance level.

(i) State the null and alternative hypotheses. **[1]**

(ii) Complete Table 6.2 and carry out the test, showing that the result is significant. **[7]**

6 (i)

6 (ii)

Expected frequency, f_e	Low	Medium	High	Total
Islands	5.2	8		**20**
Mainland countries				**30**
Total	**13**	**20**	**17**	**50**

Table 6.2

Dipali is encouraged by this result and decides to continue her investigation using all the countries covered by the pre-release data, except for the 17 countries for which the life expectancy is not given. She starts by working out the means of the life expectancies for the 68 islands and for the mainland countries.

(iii) The sum of the life expectancies for the islands is 5149.36 years.
The sum of the life expectancies for the mainland countries is 10 781.86 years.

Calculate the mean of the life expectancies for the islands and the mean of the life expectancies for the mainland countries. **[2]**

6 (iii)	

(iv) Dipali then works out the equivalent weighted means, taking the populations of the countries into account. Table 6.3 contains relevant parts of a spreadsheet to work out the weighted mean of the life expectancies of three islands. The answer for the weighted mean should appear in cell T25.

Fill in the missing numbers in Table 6.3. **[4]**

6 (iv)

	P	Q	R	T
20	**Island**	**Population**	**Life expectancy**	**Q × R**
21	Barbados	289 690	74.99	
22	Comoros	766 875	63.48	
23	New Zealand	4 401 916	80.93	
24	**Total**			
25				

Table 6.3

Dipali uses her spreadsheet correctly to find that the weighted mean life expectancy for all 68 islands is 75.62 years and that for all the mainland countries is 70.65 years.

(v) Dipali has used three different techniques in her investigation:

- a χ^2 test;
- comparing the simple means of the life expectancies of the two groups;
- comparing the weighted means.

State which of these techniques you consider to be the most appropriate and give a reason why it is better than each of the other two. The two reasons which you give should be different. **[3]**

6 (v)	

END OF QUESTION PAPER

ADDITIONAL ANSWER SPACE

If additional space is required, you should use the following lined page(s). The question number(s) must be clearly shown in the margin(s).