OCR
Oxford Cambridge and RSA

Accredited

AS Level Further Mathematics B (MEI)
Y416 Statistics b
Sample Question Paper

Date – Morning/Afternoon
Time allowed: 1 hour 15 minutes

MODEL ANSWERS

OCR supplied materials:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)
- Scientific or graphical calculator

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.**
- Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION
- The total number of marks for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- 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 used. You should communicate your method with correct reasoning.
- The Printed Answer Booklet consists of **12** pages. The Question Paper consists of **8** pages.

Answer **all** the questions.

1 Abby runs a stall at a charity event. Visitors to the stall pay to play a game in which six fair dice are rolled. If the difference between the highest and lowest scores is less than 3 then the player wins £5. Otherwise the player wins nothing.

Abby designs the spreadsheet shown in Fig. 1 to estimate the probability of a player winning, by simulating 20 goes at the game. Cell C5, highlighted, shows that the 2nd dice in simulated game 4 scores 5. Cells H5 and I5 show the highest and lowest scores, respectively, in game 4, and cell J5 gives the difference between them.

C5 | =RANDBETWEEN(1,6)

	A	B	C	D	E	F	G	H	I	J	
1		dice 1	dice 2	dice 3	dice 4	dice 5	dice 6	High score	Low score	Difference	
2	game 1	2	2	4	2	3	3	4	2	2	<3
3	game 2	2	6	3	2	1	2	6	1	5	
4	game 3	3	1	5	3	4	6	6	1	5	
5	game 4	6	5	2	5	6	3	6	2	4	
6	game 5	6	3	3	5	3	2	6	2	4	
7	game 6	5	6	3	5	1	4	6	1	5	
8	game 7	2	3	1	2	6	4	6	1	5	
9	game 8	6	6	6	6	1	5	6	1	5	
10	game 9	3	6	2	5	4	1	6	1	5	
11	game 10	5	1	1	4	6	1	6	1	5	
12	game 11	2	5	6	1	6	5	6	1	5	
13	game 12	2	5	6	6	6	6	6	2	4	
14	game 13	2	2	2	2	4	4	4	2	2	<3
15	game 14	1	6	6	6	3	5	6	1	5	
16	game 15	2	2	3	3	5	1	5	1	4	
17	game 16	1	2	3	4	3	3	4	1	3	
18	game 17	5	2	4	2	1	6	6	1	5	
19	game 18	6	1	5	2	1	5	6	1	5	
20	game 19	1	3	5	1	3	5	5	1	4	
21	game 20	5	4	3	2	5	1	[redacted]	[redacted]	[redacted]	

Fig. 1

(i) (*A*) Write down the numbers in columns H, I and J for game 20. **[1]**

(*B*) Use the spreadsheet to estimate the probability of a player winning a game. **[2]**

(ii) State how the estimate of probability in **(i)** (*B*) could be improved. **[1]**

(iii) Give one advantage and one disadvantage of using this simulation technique compared with working out the theoretical probability. **[2]**

All profit made by the stall is given to charity. Abby has to decide how much to charge players to play.

(iv) If Abby charges £1 per game, estimate the total profit when 50 players each play the game once. **[3]**

i. A) 5, 1, 4

B) estimate of $P(<3) = \frac{2}{20} = 0.1$

ii. simulate more trials of the game

iii. advantage: probabilities difficult to calculate theoretically, easier to do with the simulation

disadvantage: simulation doesn't give an exact answer

iv. profit expected $= 50 \times £1 - 50 \times 0.1 \times £5$

$= £25$

2 The cumulative distribution function of the continuous random variable, Y, is given below.

$$F(y)=\begin{cases}0 & y<0\\ \dfrac{y^3-y^2}{4} & 1\le y\le 2\\ 1 & y>2\end{cases}$$

(i) Find $P(Y\le 1.5)$ [2]

(ii) Verify that the median of Y lies between 1.6 and 1.7. [3]

(iii) Find the probability density function of Y. [2]

i. $1<1.5<2 \Rightarrow$ use $\dfrac{y^3-y^2}{4}$

$$\frac{1.5^3-1.5^2}{4}=\frac{9}{32}=0.28125$$

ii. $F(1.6)=0.384$

$F(1.7)=0.50575$

for the median, m, $F(m)=0.5$

$F(1.6)<0.5<F(1.7)$ $\therefore$ median must be between 1.6 & 1.7

$\dfrac{dF}{dy}$

iii. $f(y)=\begin{cases}\dfrac{3y^2-2y}{4} & \text{for } 1\le y\le 2\\ 0 & \text{otherwise}\end{cases}$

3 At a factory, flour is packed into bags. A model for the mass in grams of flour packed into each bag is $1500 + X$, where X is a continuous random variable with probability density function

$$f(x) = \begin{cases} kx(6-x) & 0 \le x \le 6 \\ 0 & \text{elsewhere,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{36}$. **[2]**

(ii) Find the probability that a randomly selected bag of flour contains 1505 grams of flour or more. **[3]**

(iii) Find

- the mean of X,
- the standard deviation of X. **[6]**

i. $k\int_0^6 x(6-x)\,dx = 1$

$k\left[3x^2 - \frac{1}{3}x^3\right]_0^6 = 1$

$\Rightarrow 36k = 1$

$k = \frac{1}{36}$

ii. $P(\text{Weight} \ge 1505) = P(X \ge 5)$

$= \frac{1}{36}\int_5^6 x(6-x)\,dx$

$= \frac{1}{36}\int_5^6 6x - x^2\,dx$

$= \frac{1}{36}\left[3x^2 - \frac{x^3}{3}\right]_5^6$

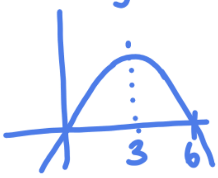
$= 0.0741$

iii. $E(X) = \int_0^6 x^2(6-x)\,dx = \int_0^6 6x^2 - x^3\,dx = \left[2x^3 - \frac{x^4}{4}\right]_0^6$

$= 3$ by calculator

Can also argue: distribution is symmetric about $x=3$

$\therefore E(X)=3$



$E(X^2) = \int_0^6 x^3(6-x)\,dx =$

$Var(X) = 10.8 - 3^2 = 1.8$

$s.d. = \sqrt{1.8} = 1.34...$

4 An online encyclopedia claims that the average mass of an adult European hedgehog is 720 g. In an investigation to check this average figure, the masses in grams of twelve randomly chosen adult European hedgehogs are measured and shown below.

705	730	720	691	718	680
731	723	745	708	724	736

(i) What assumption is required to carry out a Wilcoxon test in this situation? **[1]**

(ii) Given that this assumption is met, carry out a 2-tail Wilcoxon test at the 5% level to test whether the median mass is 720 g. You should state your hypotheses and complete the table of calculations in the Printed Answer Booklet. **[7]**

i. We assume the underlying distribution is symmetrical

ii. H_0: population median is 720g

H_1: population median is NOT 720g

mass	mass−720	mod	rank
705	−15	15	7
730	10	10	4
720	0	0	
691	−29	29	10
718	−2	2	1
680	−40	40	11
731	11	11	5
723	3	3	2
745	25	25	9
708	−12	12	6
724	4	4	3
736	16	16	8

Ignore values with mass−270=0 so n=11 not 12

$W_+ = 4+5+2+9+3+8 = 31$

$W_- = 7+10+1+11+6 = 35$

test stat : 31 ← Test stat is lowest of W^+ and W^-

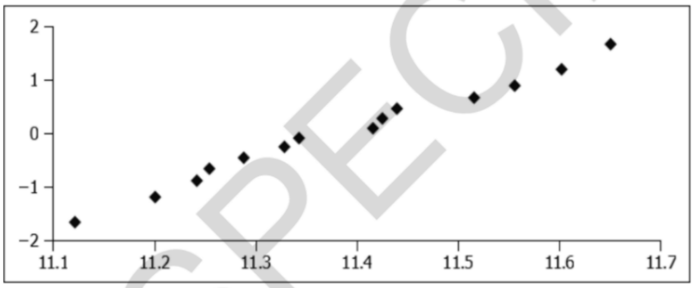
critical value : 10 ← for 2 tail 5% and $n = 11$

no evidence to reject H_0 that the median mass of hedgehogs is 750g

5 A particular alloy of bronze is specified as containing 11.5% copper on average. A researcher takes a random sample of 14 specimens of this bronze and undertakes an analysis of each of them. The percentages of copper are found to be as follows.

11.12	11.29	11.42	11.43	11.20	11.25	11.65
11.33	11.56	11.34	11.44	11.24	11.60	11.52

The researcher uses software to draw a Normal probability plot for these data and to conduct a Kolmogorov-Smirnov test for Normality. The output is shown in Fig 5.1.



Kolmogorov-Smirnov Test for Normality

p-value > 0.15

Fig 5.1

(i) Comment on what the Normal probability plot and the p-value of the test suggest about the data. **[3]**

The researcher uses software to produce a 99% confidence interval for the mean percentage of copper in the alloy, based on the t distribution. The output from the software is shown in Fig 5.2.

T Estimate of a Mean

Confidence Level 0.99

Sample

Mean 11.385

s 0.1592

N 14

Result

T Estimate of a Mean

Mean	11.385
s	0.1592
SE	0.0425
N	14
df	13
Lower Limit	11.2568
Upper Limit	11.5132
Interval	11.385 ± 0.1282

Fig 5.2

(ii) State the confidence interval which the software gives, in the form $a < \mu < b$. **[1]**

(iii) (*A*) State an assumption necessary for the use of the t distribution in the construction of this confidence interval. **[1]**

(*B*) State whether the assumption in part **(iii)** (*A*) seems reasonable. **[1]**

(iv) Does the confidence interval suggest that the copper content is different from 11.5%, on average? Explain your answer. **[2]**

(v) In the output from the software shown in Fig 5.2, SE stands for 'standard error'.

(*A*) Explain what a standard error is. **[1]**

(*B*) Show how the standard error was calculated in this case. **[1]**

(vi) Suggest a way in which the researcher could produce a narrower confidence interval. **[1]**

i. normal probability plot is approximately a straight line & high p-value ⇒ these suggest the data may be from a Normal distribution

ii. $11.2568 < \mu < 11.5132$

iii. A) we assume the underlying distribution is Normally distributed

B) result in (i) supports this assumption

iv. the confidence interval doesn't imply the mean differs from 11.5%, as 11.5 is within the interval

v. A) standard error is the standard deviation of the sample mean

B) $\dfrac{0.1592}{\sqrt{14}} = 0.0425$ s $\sqrt{N}$

vi. increase sample size (N) or use lower confidence level

6 The table below shows the mean and variance of the test scores of a random samples of 70 girls who are starting an A level Mathematics course.

Sample mean	Sample variance
118.86	86.57

(i) Showing your working, find a 95% confidence interval for the population mean. **[4]**

(ii) Explain why you can construct the interval in part **(i)** despite no information about the distribution of the parent population being given. **[2]**

(iii) The same random sample of girls repeats the test. The mean improvement in score is 0.9. The 95% confidence interval for the improvement is [-1.5, 3.3]. What is the sample variance for the improvement in score? **[2]**

i. CI given by $118.86 \pm 1.96 \times \sqrt{\frac{86.57}{70}}$

$\rightarrow 116.68 \leq \mu \leq 121.04$

ii. we have a large sample & the Central Limit Theorem states that sample means are roughly Normally distributed

iii. CI width $= 2 \times 1.96 \times \sqrt{\frac{s^2}{70}} = 4.8$

$\sqrt{\frac{s^2}{70}} = \frac{60}{49} \Longrightarrow s^2 = 104.96$

7 Two flatmates work at the same location. One of them takes the bus to work and the other one cycles. Journey times, measured in minutes, are distributed as follows.

- By bus: Normally distributed with mean 23 and standard deviation 6
- By bicycle: Normally distributed with mean 21 and standard deviation 2

You should assume that all journey times are independent.

(i) One morning the two flatmates set out at the same time. Find the probability that the person who takes the bus arrives before the cyclist. **[3]**

(ii) Find the probability that the total time taken for 5 bus journeys is less than 2 hours. **[2]**

(iii) Comment on the assumption that all journey times are independent. **[1]**

END OF QUESTION PAPER

i. bus time $= X$ cycle time $= Y$

$\bar{X} - \bar{Y} = 2$ $\sigma_x^2 + \sigma_y^2 = 40$

$\rightarrow X - Y \sim N(2, 40)$

$P(|X - Y| < 0) = 0.376$

ii. $5X \sim N(115, 180)$

$P(5X < 2\text{hrs}) = 0.645$

iii. this assumption won't always be valid, as there are factors (e.g. road works) that can delay both journeys on consecutive days.