OCR
Oxford Cambridge and RSA

AS Level Further Mathematics B (MEI)

Y416/01 Statistics b
Question Paper

Friday 22 June 2018 – Morning
Time allowed: 1 hour 15 minutes

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You may use:
- a scientific or graphical calculator

MODEL ANSWERS

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.** If additional space is required, you should use the lined page(s) at the end of the Printed Answer Booklet. The question number(s) must be clearly shown.
- Do **not** write in the barcodes.
- 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.

1 The birth weights, in kilograms, of a random sample of 9 captive-bred elephants are as follows.

94	138	130	118	146	165	82	115	69

A researcher uses software to produce a 99% confidence interval for the mean birth weight of captive-bred elephants. The output from the software is shown in Fig. 1.

T Estimate of a Mean

Confidence Level 0.99

Sample

Mean 117.444

s 31.2574

N 9

Result

T Estimate of a Mean

Mean	117.444
s	31.2574
SE	10.4191
N	9
df	8
Lower limit	82.4838
Upper limit	152.4042
Interval	117.444 ± 34.9602

Fig. 1

(i) State an assumption about the distribution of the population from which these weights come that is necessary in order to produce this interval. **[1]**

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

(iii) Explain
- what the label df means,
- how the value of df is calculated for a confidence interval produced using the t distribution. **[2]**

(iv) State two ways in which the researcher could have obtained a narrower confidence interval. **[2]**

i. the underlying distribution must be Normal

ii. $82.4838 < \mu < 152.4042$

iii. df: degrees of freedom. its value is calculated as number of data values -1 (or -2 for a 2-sample)

iv. increase sample size & lower confidence level

2 A supermarket sells oranges. Their weights are modelled by the random variable X which has a Normal distribution with mean 345 grams and standard deviation 15 grams. When the oranges have been peeled, their weights in grams, Y, are modelled by $Y = 0.7X$.

(i) Find the probability that a randomly chosen peeled orange weighs less than 250 grams. **[3]**

I randomly choose 5 oranges to buy.

(ii) Find the probability that the total weight of the 5 unpeeled oranges is at least 1800 grams. **[2]**

(iii) I peel three of the oranges and leave the remaining two unpeeled. Find the probability that the total weight of the two unpeeled oranges is greater than the total weight of the three peeled ones. **[4]**

i. $Y \sim N(0.7 \times 345, (0.7 \times 15)^2)$ square it as SD given

$\rightarrow Y \sim N(241.5, 110.25)$

$P(Y < 250) = 0.791$

ii. $5Y \sim N(1725, 1125)$

$P(Y \geq 1800) = 0.0127$

'at least'

iii. Mean $= 2 \times 345 - 3 \times 241.5 = -34.5$

Variance $= 2 \times 15^2 + 3 \times 10.5^2 = 780.75$

$(2X - 3Y) \sim N(-34.5, 780.75)$

$P((2X - 3Y) > 0) = 0.108$

3 The probability density function of the continuous random variable X is given by

$$f(x) = \begin{cases} c+x & -c \leqslant x \leqslant 0, \\ c-x & 0 \leqslant x \leqslant c, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a positive constant.

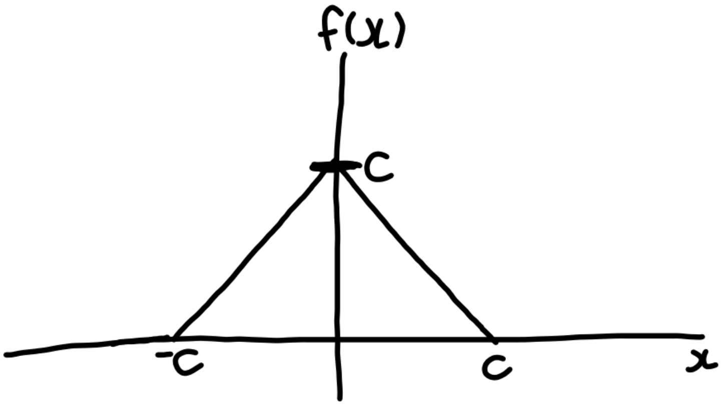
(i) (A) Sketch the graph of the probability density function. **[2]**

(B) Show that $c = 1$. **[2]**

(ii) Find $\mathrm{P}\left(X < \frac{1}{4}\right)$. **[2]**

(iii) Find
- the mean of X,
- the standard deviation of X. **[4]**

i. A)



B) $\frac{1}{2} \times 2c \times c = 1$ (area of triangle)

$c^2 = 1$

$\Rightarrow c = 1 \quad (c > 0)$

ii. $P(X < \frac{1}{4}) = \frac{1}{2} + \left(\frac{1+0.75}{2}\right) \times 0.25 = 0.71875 \, / \, \frac{23}{32}$

P(X<0)

iii. $E(X) = 0$ by symmetry

$E(X^2) = \int_{-1}^{0} (x^2 + x^3)\,dx + \int_0^1 (x^2 - x^3)\,dx = \left[\frac{1}{3}x^3 + \frac{1}{4}x^4\right]_{-1}^{0} + \left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_0^1$

$= \frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4} = \frac{2}{12} = \frac{1}{6} \quad \therefore \mathrm{Var}(X) = \frac{1}{6}$

$\text{s.d.} = \sqrt{\frac{1}{6}} = 0.408$

4 The random variable X has a continuous uniform distribution on [0, 10].

(i) Find $\mathrm{P}(3 < X < 6)$. **[1]**

(ii) Find each of the following.
- $\mathrm{E}(X)$
- $\mathrm{Var}(X)$

[2]

Marisa is investigating the sample mean, Y, of 8 independent values of X. She designs a simulation shown in the spreadsheet in Fig. 4.1. Each of the 25 rows below the heading row consists of 8 values of X together with the value of Y. All of the values in the spreadsheet have been rounded to 2 decimal places.

	A	B	C	D	E	F	G	H	I	J
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y	
2	6.31	2.45	3.27	3.06	4.16	1.53	0.43	7.99	3.65	
3	1.70	1.52	7.10	8.93	6.44	2.70	9.96	7.83	5.77	
4	9.15	0.52	4.95	6.99	6.52	3.15	0.81	5.35	4.68	
5	0.65	2.71	7.92	9.65	0.50	4.87	6.46	2.67	4.43	
6	3.09	6.11	3.96	0.09	0.18	4.67	0.67	6.20	3.12	
7	7.06	5.84	1.97	3.60	9.36	1.97	4.48	3.47	4.72	
8	1.46	1.57	5.45	0.37	3.76	7.56	8.48	9.12	4.72	
9	9.42	1.85	4.91	1.61	1.94	8.00	1.77	5.34	4.36	
10	2.98	5.32	2.91	4.12	9.16	1.76	9.97	6.88	5.39	
11	2.83	3.44	3.28	7.85	1.00	0.93	8.77	4.03	4.01	
12	4.51	0.59	5.84	9.87	8.65	3.94	7.18	0.23	5.10	
13	4.49	0.69	3.65	8.78	4.96	8.96	3.77	1.43	4.59	
14	6.57	8.08	4.85	6.75	7.92	0.27	9.69	4.04	6.02	← >6
15	8.35	1.09	8.63	8.04	7.23	2.12	2.57	9.59	5.95	
16	5.24	9.53	6.08	8.21	3.61	7.07	6.65	7.63	6.75	← >6
17	7.89	5.50	3.09	0.71	6.47	5.49	6.47	4.95	5.07	
18	8.36	7.27	2.35	9.04	0.58	2.26	3.01	7.90	5.10	
19	3.76	1.01	9.61	9.65	7.89	9.98	6.28	4.34	6.56	← >6
20	9.94	6.84	3.38	5.53	0.26	8.53	5.72	5.12	5.66	
21	7.25	9.10	0.34	2.88	4.66	2.65	6.37	7.63	5.11	
22	7.18	7.14	5.38	0.04	4.09	6.47	4.96	4.23	4.94	
23	8.69	5.04	4.90	2.94	2.00	4.23	4.13	0.97	4.11	
24	3.46	6.33	0.48	9.35	0.23	1.18	7.97	6.37	4.42	
25	2.37	7.26	7.16	1.24	5.26	2.80	3.55	3.84	4.19	
26	2.16	8.30	7.17	3.32	2.96	1.30	9.11	0.31	4.33	
27										

Fig. 4.1

(iii) Use the spreadsheet to estimate $\mathrm{P}(3 < Y < 6)$. **[2]**

(iv) Explain why it is not surprising that this estimated probability is substantially greater than the value which you calculated in part **(i)**. **[2]**

Marisa wonders whether, even though the sample size is only 8, use of the Central Limit Theorem will provide a good approximation to $\mathrm{P}(3 < Y < 6)$.

(v) Calculate an estimate of $\mathrm{P}(3 < Y < 6)$ using the Central Limit Theorem. **[3]**

A Normal probability plot of the 25 simulated values of Y is shown in Fig. 4.2.

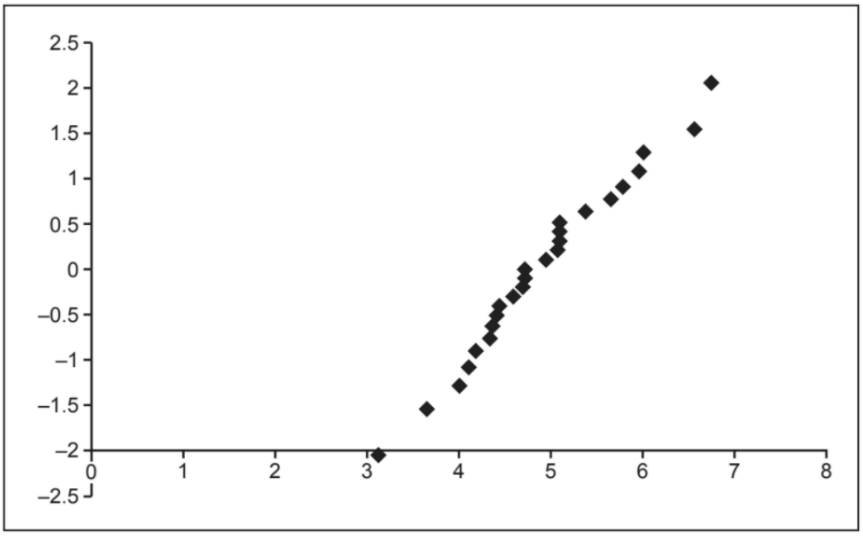


Fig. 4.2

(vi) Explain what the Normal probability plot suggests about the use of the Central Limit Theorem to approximate $P(3 < Y < 6)$. **[2]**

Marisa now decides to use a spreadsheet with 1000 rows below the heading row, rather than the 25 which she used in the initial simulation shown in Fig. 4.1. She uses a counter to count the number of values of Y between 3 and 6. This value is 808.

(vii) Explain whether the value 808 supports the suggestion that the Central Limit Theorem provides a good approximation to $P(3 < Y < 6)$. **[1]**

Marisa decides to repeat each of her two simulations many times in order to investigate how variable the probability estimates are in each case.

(viii) Explain whether you would expect there to be more, the same or less variability in the probability estimates based on 1000 rows than in the probability estimates based on 25 rows. **[2]**

i. $P(3<X<6) = 0.3$ (height is $\frac{1}{10}$)

ii. $E(X) = 5$ $\quad E(X^2) = \frac{1}{10}\int_0^{10} x^2\,dx = \frac{1}{10}\left(\frac{10^3}{3}\right)$
$= \frac{100}{3}$

$Var(X) = \frac{100}{3} - 25$
$= \frac{25}{3}$

iii. 3 points outside range $\Rightarrow P(3<Y<6) = \frac{22}{25} = 0.88$

iv. because the mean of Y = mean of X, but the Variance of Y is only $\frac{1}{8} \times Var(X)$

v. using the CLT: can approximate dist. of Y as

$Y \sim N(5, \frac{25}{24})$

$\Rightarrow P(3 < Y < 6) = 0.811$ approx.

vi. Normal probability plot is very close to a straight line. This suggests the dist. may be Normal & $\therefore$ CLT gives a good approximation

vii. Yes – the estimated probability from $\frac{808}{1000} = 0.808$. This is $\simeq 0.811$ from the CLT.

viii. We expect less variation with 1000 rows because increasing the number of rows means the estimated probability tends closer towards the true value

5 The flight time between two airports is known to be Normally distributed with mean 3.75 hours and standard deviation 0.21 hours. A new airline starts flying the same route. The flight times for a random sample of 12 flights with the new airline are shown in the spreadsheet (Fig. 5), together with the sample mean.

	A	B	C	D	E	F	G	H	I	J	K	L
1	3.595	3.723	3.584	3.643	3.669	3.697	3.550	3.674	3.924	3.563	3.330	3.706
2												
3	Mean	3.638										

Fig. 5

(i) In this question you must show detailed reasoning.

You should assume that:
- the flight times for the new airline are Normally distributed,
- the standard deviation of the flight times is still 0.21 hours.

Carry out a test at the 5% significance level to investigate whether the mean flight time for the new airline is less than 3.75 hours. **[7]**

(ii) If both of the assumptions in part **(i)** were false, name an alternative test that you could carry out to investigate average flight times, stating any assumption necessary for this test. **[2]**

(iii) If instead the flight times were still Normally distributed but the standard deviation was not known to be 0.21 hours, name another test that you could carry out. **[1]**

i. $H_0: \mu = 3.75 \quad H_1: \mu < 3.75$

where μ = population mean flight time with new airline (hours)

$$\text{test stat.} = \frac{3.638 - 3.75}{0.21/\sqrt{12}} = -1.848 \quad \left(\frac{\bar{x} - \mu}{s/\sqrt{n}}\right)$$

for one-tailed test: critical value @ 5% level is -1.645. $-1.848 < -1.645$ $\therefore$ significant value, reject H_0. there is sufficient evidence to suggest the average flight time is < 3.75 hrs with the new airline.

ii. Wilcoxon signed rank test $\Rightarrow$ underlying dist. must be symmetrical.

iii. single sample t-test

6 A company has a large fleet of cars. It is claimed that use of a fuel additive will reduce fuel consumption. In order to test this claim a researcher at the company randomly selects 40 of the cars. The fuel consumption of each of the cars is measured, both with and without the fuel additive. The researcher then calculates the difference d litres per kilometre between the two figures for each car, where d is the fuel consumption without the additive minus the fuel consumption with the additive. The sample mean of d is 0.29 and the sample standard deviation is 1.64.

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

(ii) Explain whether the confidence interval suggests that, on average, the fuel additive does reduce fuel consumption. **[2]**

(iii) Explain why you can construct the interval in part **(i)** despite not having any information about the distribution of the population of differences. **[2]**

(iv) Explain why the sample used was random. **[2]**

END OF QUESTION PAPER

i. CI given by $0.29 \pm 1.96 \times \frac{1.64}{\sqrt{40}}$

$\Rightarrow -0.218 \leq \mu \leq 0.798$

ii. this confidence interval contains 0, so doesn't suggest additive reduces fuel consumption

iii. We can use that the sample is large to apply the Central Limit Theorem, which states that sample means are thus approximately Normally distributed.

iv. a random sample enables proper inference about the population to be undertaken.