OCR
Oxford Cambridge and RSA

Thursday 6 June 2019 – Afternoon

AS Level Further Mathematics B (MEI)

Y416/01 Statistics b

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.
- 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.
- 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 It is known that the red blood cell count of adults in a particular country, measured in suitable units, has mean 4.96 and variance 0.15.

(a) Find the probability that the mean red blood cell count of a random sample of 50 adults from this country is at least 5.00. [3]

(b) Explain how you can find the probability in part (a) despite the fact that you do not know the distribution of red blood cell counts. [3]

a) central limit theorem: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

distribution is $\sim N(4.96, \frac{0.15}{50})$

$P(\text{mean} > 5.00) = 0.233$ by calculator

b) by the central limit theorem, for large values of n (e.g. n = 50), the distribution of the sample mean is approximately Normal

2 Leila and Caleb are playing a game, using fair six-sided dice and unbiased coins.

- Leila rolls two dice, and her score L is the total of the scores on the two dice.
- Caleb spins 4 coins and his score C is three times the number of heads obtained.

The winner of a game is the player with the higher score. If the two scores are equal, the result of the game is a draw. The spreadsheet in Fig. 2 shows a simulation of 20 plays of the game.

	A	B	C	D	E	F	G	H	
1	**First dice**	**Second dice**	**Total (Leila's score) *L***	**Coin 1**	**Coin 2**	**Coin 3**	**Coin 4**	**Caleb's score *C***	
2	1	2	3	H	T	T	T	3	
3	6	1	7 b	T	H	T	T	3	
4	2	6	8 b	H	H	T	T	6	
5	2	5	7 b	T	H	H	H	9 b	c
6	1	5	6	T	H	T	T	3	
7	5	2	7 b	H	H	H	H	12 b	c
8	1	1	2	H	T	H	T	6	c
9	2	6	8 b	T	H	T	H	6	
10	6	2	8 b	H	T	H	T	6	
11	1	3	4	T	H	H	H	9 b	c
12	6	1	7 b	T	H	T	T	3	
13	3	1	4	T	T	T	T	0	
14	3	6	9 b	H	T	H	H	9 b	
15	2	3	5	T	H	H	H	9 b	c
16	2	5	7 b	H	H	H	H	12 b	c
17	1	5	6	H	H	T	H	9 b	c
18	5	6	11 b	T	H	H	H	9 b	
19	4	2	6	T	H	H	T	6	
20	6	5	11 b	T	T	H	H	6	
21	1	1	2	T	T	T	T	0	

Fig. 2

(a) Explain why the value of C in row 2 is 3. **[1]**

(b) Use the spreadsheet to estimate $P(C > 6)$ and $P(L > 6)$. **[2]**

(c) Use the spreadsheet to estimate the probability that Leila loses a randomly chosen game. **[2]**

(d) Explain why your answers to parts **(b)** and **(c)** may not be very close to the true values. **[1]**

(e) Leila claims that the game is fair (that Leila and Caleb each have an equal chance of winning) because both she and Caleb can get a maximum score of 12 and also in the simulation she won exactly 50% of the games.
Make 2 comments about Leila's claim. **[2]**

a) because there is one head & the score = 3 x the number of heads

b) instances where L & C > 6 marked 'b'

estimate of $P(C>6) = \frac{8}{20} = 0.4$

estimate of $P(L>6) = \frac{11}{20} = 0.55$

c) instances where L loses marked 'c'

$\Rightarrow P(L \text{ loses}) = \frac{7}{20} = 0.35$

d) because the number of simulations is small

e) there were some draws, so C didn't win 50% of games $\Rightarrow$ L's claim may not be true.

while the max. score is the same, L's minimum is 2 whereas C's is 0, so claim isn't valid for only comparing max. values.

other things you could comment on:

- larger sample size needed for confidence in claim
- expected scores

3 A bus runs from point A on the outskirts of a city, stops at point B outside the rail station, and continues to point C in the city centre.
The journey times for the sections A to B and B to C vary according to traffic conditions, and are modelled by independent Normal distributions with means and standard deviations as shown in the table.

	Journey time (minutes)	
	Mean	Standard deviation
A to B	21	3
B to C	29	4

(a) Find the probability that a randomly chosen journey from A to B takes less than the scheduled time of 23 minutes. **[1]**

For every journey, the bus stops for 1 minute when it reaches B to drop off and pick up passengers.

(b) Find the probability that a randomly chosen journey from A to C takes less than the scheduled time of 50 minutes. **[4]**

Mary travels on the bus from the station at B to her workplace at C every working day. You should assume that times for her bus journeys on different days are independent.

(c) Find the probability that the total time taken for her five journeys on the bus in a randomly chosen week is at least $2\frac{1}{2}$ hours. **[3]**

(d) Comment on the assumption that times on different days are independent. **[1]**

a) $P(t_{AB}<23) = 0.748$ by calculator

b) extra +1 min not included in journey time, so need to account for it

3^2+4^2

for t_{AC}, add μ_{AB} & μ_{BC}, σ_{AB}^2 & σ_{BC}^2 : $t_{AC} \sim N(50, 25)+1$

$P(\underline{\text{total}}\text{ time} < 50) = P(t_{AC} < 49) = 0.421$ by calculator

c) 5 journeys from B to C : $5t_{BC} \sim N(5\times29, 5\times4^2)$

$\sim N(145, 80)$

$P(5t_{BC} > 150) = 0.288$ by calculator

2.5 hrs

d) it seems likely that the assumption is valid, since it is unlikely that a delay on one day would affect another delay

this is a question where you can argue either way as long as you justify your answer. e.g. you could argue it isn't valid because there may be bad weather or road works delaying traffic for a week

4 The cumulative distribution function of the continuous random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0, \\ k(12x - x^2) & 0 \leqslant x \leqslant 2, \\ 1 & x > 2, \end{cases}$$

where k is a constant.

(a) Show that $k = 0.05$. **[2]**

(b) Find $P(1 \leqslant X \leqslant 1.5)$. **[2]**

(c) Find the median of X, correct to 3 significant figures. **[3]**

(d) Find which of the median, mean and mode of X is the largest of the three measures of central tendency. **[5]**

a) continuous $\Rightarrow$ no jump @ $x = 2$

so $F(2) = 1 = k(12(2) - 2^2)$

$(24 - 4)k = 1$

$20k = 1$

$\therefore k = 0.05$

b) $P(1 \leq X \leq 1.5) = F(1.5) - F(1)$

$= 0.05(12(1.5) - 1.5^2) - 0.05(12(1) - 1^2)$

$= 0.7875 - 0.55$

$= 0.2375$

c) median $\Rightarrow$ cumulative probability $= 0.5$

$0.05(12m - m^2) = 0.5$

$0.05m^2 - 0.6m + 0.5 = 0$

$m = \dfrac{0.6 \pm \sqrt{0.6^2 - 4(0.05)(0.5)}}{2(0.05)} = 11.099$ or 0.901

11.099 outside range so reject $\therefore$ $M = 0.901$

d) for mean & mode need $f(x) = F'(x)$

$f(x) = 0.05(12 - 2x)$, $0 \le x \le 2$

$E(X) = \int_0^2 0.05x(12-2x)\,dx = 0.05\left[6x^2 - \frac{2}{3}x^3\right]_0^2$

$= 0.933$

small value subtracted

$f(x)$ largest when x smallest (in range $[0,2]$) so max. value

$= 0 =$ mode

$0.933 > 0.901 > 0$ so mean is largest

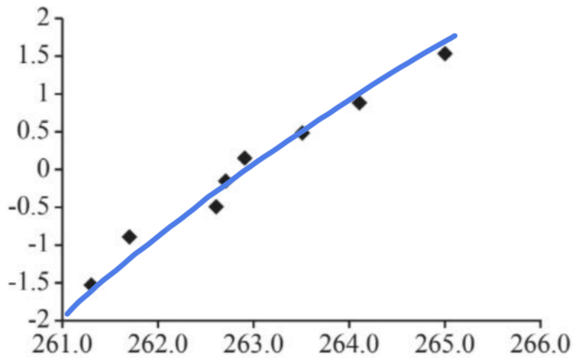
5 A technician is investigating whether a batch of nylon 66 (a particular type of nylon) is contaminated by another type of nylon.
The average melting point of nylon 66 is 264 °C. However, if the batch is contaminated by the other type of nylon the melting point will be lower. The melting points, in °C, of a random sample of 8 pieces of nylon from the batch are as follows.

262.7	265.0	264.1	261.7	262.9	263.5	261.3	262.6

(a) Find

- the sample mean,
- the sample standard deviation. **[2]**

The technician produces a Normal probability plot and carries out a Kolmogorov-Smirnov test for these data as shown in Fig. 5.



Kolmogorov-Smirnov test for Normality

p-value = 0.86

Fig. 5

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

(c) **In this question you must show detailed reasoning.**

Carry out a suitable test at the 5% significance level to investigate whether the batch appears to be contaminated with another type of nylon. **[8]**

(d) Name an alternative test that could have been carried out if the population standard deviation had been known. **[1]**

a) by calculator:

sample mean = 262.975

sample standard deviation = 1.213

b) the Normal probability plot is roughly straight & we have a very high p-value, suggesting the data may be

Normally distributed.

c) 1. state hypotheses: $H_0: \mu = 264$ $H_1: \mu < 264$ remember to define your variables in context

where μ = population mean melting temperature

2. find test statistic

t-test: $t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{262.975 - 264}{1.213/\sqrt{8}} = -2.391$

3. find critical value & draw conclusions

7 d.o.f. $\Rightarrow t_7$, need c.v. for d.o.f. = 7, significance level 0.05

(one-tailed): C.V. = 1.895

$-2.391 < -1.895$, ∴ significant, can reject H_0.

there is sufficient evidence to suggest that the batch might be contaminated with another type of nylon ← link back to context

d) a test based on the Normal distribution

↳ z-test

6 The label on a pack of strawberries in a large batch states that it holds 250 g of strawberries.
A random sample of 40 packs from the batch is selected and software is used to produce a 95% confidence interval for the mean weight of strawberries per pack. An extract from the software output is shown in Fig. 6.

Sample Mean	248.92
Standard Error	0.61506
Sample Size	40
Confidence Level	0.95
Interval	248.92 ± 1.2055

Fig. 6

(a) Explain whether the confidence interval suggests that the mean weight of strawberries per pack in the batch is different from 250 g. **[2]**

(b) A manager looking at the data says that the conclusion would have been different if a 90% confidence interval had been used.
Determine whether the manager is correct. **[3]**

(c) Explain briefly whether or not it is appropriate for the manager to vary the confidence level before coming to any conclusions. **[2]**

(d) On another occasion, using the same sample size, a 95% confidence interval for the mean weight of strawberries per pack is [248.05, 249.95].
Find the sample variance in this case. **[2]**

(e) Explain the meaning of a 95% confidence interval. **[2]**

END OF QUESTION PAPER

a) the confidence level does not suggest the mean weight is different from 250g $\because$ the upper bound is 250.1255 $\Rightarrow$ the interval contains 250

b) z-score for 90% confidence level is 1.645

so interval given by $\bar{x} \pm z \times$ standard error

$\hookrightarrow 248.92 \pm 1.645 \times 0.61506$

248.92 ± 1.0118

$247.91 < \mu < 249.93 \Rightarrow$ excludes 250 so manager is correct

c) no, it is not appropriate as the size of the interval should be decided before it is calculated. otherwise, the level can be adjusted to provide the desired conclusion.

d) $249.95 - 248.05 = 1.9$ z for 95% level

width of interval $= 2 \times 1.96 \times \sqrt{\frac{s^2}{40}}$ error n

$= 1.9$

$$\sqrt{\frac{s^2}{40}} = \frac{95}{196}$$

$\therefore s^2 = 9.397 \leftarrow$ variance

e) in repeated sampling, 95% of confidence intervals constructed in this way will contain the true population mean.