Answer ALL questions. Write your answers in the spaces provided.

1. A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

(a) Describe how this sample could be taken.

(3)

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s$$

(b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

(1)

(c) Describe one limitation of this regression model.

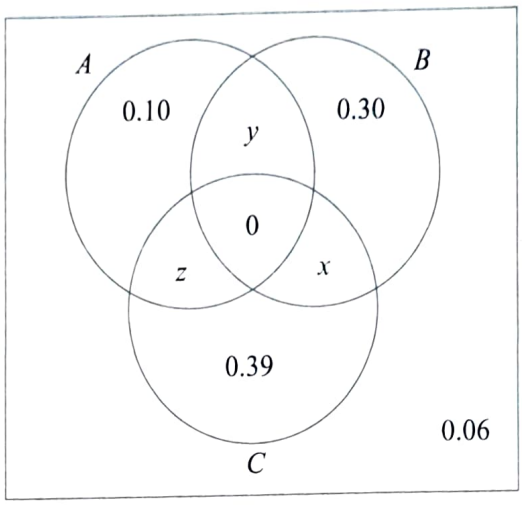
(1)

a) Label each [redacted] student by year group.
Use random numbers to [redacted] select
24 Year 12s and 16 Year 13s.

b) Increase by 2.8 marks.

c) The best performance is predicted for students who never wake up.

2. The Venn diagram shows three events, A, B and C, and their associated probabilities.



Events B and C are mutually exclusive.
Events A and C are independent.

Showing your working, find the value of x, the value of y and the value of z.

(5)

$x = 0$.

$P(A) = 0.1 + y + z$

$P(B) = 0.3 + y$

$P(C) = 0.39 + z$.

$P(A \cap C) = z$.

$P(A) \times P(C) = P(A \cap C)$

$(0.1 + y + z)(0.39 + z) = z$

~~$= 0.039 + 0.39y + 0.39z + \ldots$~~.

$0.1 + 0.3 + 0.39 + 0.06 + y + z = 1$

$\Rightarrow y + z = 0.15$

$0.25(0.39 + z) = z$

$0.0975 = 0.75z$

$\Rightarrow z = 0.13, \quad y = 0.02$.

3. A fair 5-sided spinner has sides numbered 1, 2, 3, 4 and 5

The spinner is spun once and the score of the side it lands on is recorded.

(a) Write down the name of the distribution that can be used to model the score of the side it lands on.

(1)

The spinner is spun 28 times.

The random variable X represents the number of times the spinner lands on 2

(b) (i) Find the probability that the spinner lands on 2 at least 7 times.

(ii) Find $P(4 \leqslant X < 8)$

(5)

a) Discrete uniform.

b) i) $B(28, 0.2)$

$P(X \geqslant 7) = 1 - P(X \leqslant 6)$

$= 0.322.$

ii) $P(4 \leq X < 8) = P(X \leq 7) - P(X \leq 3)$

$= 0.658$

4. Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	Frequency	Midpoint (x mm)
$0 \leqslant r < 0.5$	121	0.25
$0.5 \leqslant r < 1.0$	10	0.75
$1.0 \leqslant r < 5.0$	24	3.0
$5.0 \leqslant r < 10.0$	12	7.5
$10.0 \leqslant r < 30.0$	17	20.0

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

a) Trace, data needs to be converted to numbers first.

b) 184 values → 138th value required.

$$1 + \frac{4}{24}(138 - 131) = 2.16\dot{6} \text{ mm}.$$

Question 4 continued

c)

$$\sigma = \sqrt{\frac{7704.1875}{184} - \left(\frac{539.75}{184}\right)^2}$$

$$= 5.7676.$$

d) i) The data is evenly spread amongst each class.

ii) The majority of values in the first class would be zero.

iii) Smaller, the sum Σfx ought to be smaller (considering many values will be zero, rather than 0.25).

5. Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief.

Julie's hypothesis test is shown below.

$H_0 : p = 0.15$

$H_1 : p \geqslant 0.15$

Let X = the number of customers who buy chocolate.

$X \sim B(30, 0.15)$

$P(X = 8) = 0.0420$

$0.0420 < 0.05$ so reject H_0

There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test.

(2)

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test. Give a reason for your answer.

(1)

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05

(2)

(d) Find the actual level of significance of this test.

(1)

a) • H_1 : $p > 0.15$.
• Needs to calculate $P(X \geqslant 8)$, not $P(X = 8)$.

b) $P(X \geqslant 8) = 0.0698 > 0.05$.

The null hypothesis H_0 would not be rejected, so yes, it will affect the conclusion.

Question 5 continued

c) $\mathbb{P}(X \leqslant 8) = 0.9722 > 0.95$.

The critical region is
$X \geqslant 9$.

d) 0.0278