

IB Maths AA SL – Topic 4: Statistics and Probability

IB AA SL · Total marks: 73 · Answers on the last page

Covers descriptive statistics, probability (Venn diagrams, tree diagrams, conditional probability), the binomial and normal distributions and linear regression. A GDC is allowed throughout; allow about 75 minutes. Give numerical answers to 3 significant figures unless exact.

Section A – Skills (Short questions. GDC allowed.)

- The number of goals scored by a hockey team in eight matches was 12, 15, 9, 18, 21, 15, 11, 24.
 - Find the mean and the median.
 - Write down the mode and the range.
 - Find the standard deviation. [5]
- The times, t minutes, taken by 40 students to complete a puzzle are grouped as follows: $0 \leq t < 10$: 4 students; $10 \leq t < 20$: 9; $20 \leq t < 30$: 15; $30 \leq t < 40$: 8; $40 \leq t < 50$: 4.
 - Write down the modal class.
 - Estimate the mean time.
 - Explain why your answer to (b) is only an estimate. [4]
- A box-and-whisker diagram for a data set has minimum 5, lower quartile 14, median 19, upper quartile 26 and maximum 40.
 - Find the interquartile range.
 - Determine, with a calculation, whether the value 40 is an outlier.
 - The maximum is later corrected to 45. State whether this changes your conclusion in (b). [4]
- Events A and B are such that $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cap B) = 0.15$.
 - Find $P(A \cup B)$.
 - Find $P(A | B)$.
 - Determine whether A and B are independent. Justify your answer. [5]
- A discrete random variable X has probability distribution $P(X = x) = kx$ for $x = 1, 2, 3, 4$.
 - Show that $k = \frac{1}{10}$.
 - Find $E(X)$.
 - Find $P(X \geq 2 | X \leq 3)$. [5]

Section B – Exam-style (Show full method. GDC allowed.)

- A bag contains 5 red and 3 blue counters. Two counters are taken at random, one after the other, without replacement.
 - Find the probability that both counters are the same colour.
 - Find the probability that the second counter is red, given that the first counter is blue.
 - Find the probability that the first counter is blue, given that the second counter is red. [6]
- A multiple-choice test has 10 questions, each with a probability 0.3 of being answered correctly by a student who guesses. Let X be the number of correct answers.
 - State the distribution of X and any assumptions required.
 - Find $P(X = 3)$.
 - Find $P(X \geq 4)$.

- (d) Find the expected value and the variance of X . [7]
8. The heights of adult males in a town are normally distributed with mean 170 cm and standard deviation 8 cm.
- (a) Find the probability that a randomly chosen man is shorter than 160 cm.
 - (b) Find the probability that a randomly chosen man is between 165 cm and 180 cm tall.
 - (c) The tallest 10 percent of men are classed as 'very tall'. Find the minimum height of a very tall man. [6]
9. The table shows the number of hours, x , six students spent on social media per week and their score, y , in a test.
 x : 2, 4, 5, 7, 8, 10; y : 65, 58, 55, 48, 44, 37.
- (a) Find Pearson's product-moment correlation coefficient r and describe the correlation.
 - (b) Find the equation of the regression line of y on x .
 - (c) Use your line to estimate the score of a student who spends 6 hours per week on social media, and comment on the reliability of this estimate.
 - (d) Explain why it would not be appropriate to use the line to estimate y when $x = 30$. [8]
10. The mass of a bag of rice is normally distributed with mean μ g and standard deviation 5 g. It is known that 20 percent of bags have mass less than 50 g.
- (a) Find μ .
 - (b) Bags with mass less than 50 g are rejected. Eight bags are chosen at random. Find the probability that at least one bag is rejected. [6]

Section C – Extended (Longer structured questions.)

11. The time taken by a courier to deliver a parcel is normally distributed with mean 68 minutes and standard deviation 6 minutes. A delivery is 'late' if it takes more than 75 minutes.
- (a) Show that the probability that a delivery is late is 0.122, correct to 3 significant figures.
 - (b) On one day the courier makes 8 deliveries. Find the probability that exactly 2 are late.
 - (c) Find the probability that at least one delivery is late.
 - (d) The courier's manager wants the probability of a delivery being late to be 0.05. Assuming the mean stays at 68 minutes, find the standard deviation required. [9]
12. A scientist records the length, x cm, of a plant and the number of leaves, y , on seven plants:
 x : 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5; y : 12, 15, 19, 22, 27, 29, 34.
- (a) Write down the coordinates of the mean point $(\bar{x}, \bar{y})$.
 - (b) Find the equation of the regression line of y on x in the form $y = ax + b$.
 - (c) Interpret the value of a in context.
 - (d) Estimate the number of leaves on a plant of length 5 cm and state one reason your estimate may be unreliable.
 - (e) The scientist claims that a longer plant causes more leaves to grow. Comment on this claim. [8]

Answers

Check your working against these. If an answer does not match, try the question again before looking for the method.

- Mean = 15.6 (15.625), median = 15
 - Mode = 15, range = 15
 - $\sigma = 4.79$
- $20 \leq t < 30$
 - $\bar{t} = \frac{990}{40} = 24.75 \approx 24.8$ minutes
 - Mid-interval values are used; the actual times within each class are unknown.
- IQR = 12
 - Upper fence = $26 + 1.5 \times 12 = 44$; $40 < 44$ so not an outlier
 - Yes: $45 > 44$, so 45 is an outlier.
- $0.4 + 0.5 - 0.15 = 0.75$
 - $\frac{0.15}{0.5} = 0.3$
 - Not independent: $P(A)P(B) = 0.2$
 $e0.15$ (or $P(A | B) = 0.3$
 $eP(A)$).
- $k(1 + 2 + 3 + 4) = 1 \Rightarrow 10k = 1$
 - $E(X) = \frac{1 + 4 + 9 + 16}{10} = 3$
 - $\frac{P(X = 2) + P(X = 3)}{P(X \leq 3)} = \frac{0.5}{0.6} = \frac{5}{6}$
- $\frac{5}{8} \cdot \frac{4}{7} + \frac{3}{8} \cdot \frac{2}{7} = \frac{26}{56} = \frac{13}{28}$
 - $\frac{5}{7}$
 - $P(2\text{nd red}) = \frac{5}{8}$, so $\frac{(3/8)(5/7)}{5/8} = \frac{3}{7}$
- $X \sim B(10, 0.3)$; each question independent, constant probability 0.3
 - 0.267
 - 0.350
 - $E(X) = 3$, $\text{Var}(X) = 2.1$
- 0.106
 - 0.628
 - 180 cm (180.25)
- $r = -1.00$ (-0.99968); very strong negative (linear) correlation
 - $y = -3.5x + 72.2$
 - $y = 51.2$; reliable since r is close to -1 and $x = 6$ lies within the data range (interpolation)
 - $x = 30$ is outside the range of the data (extrapolation); the linear trend may not continue.
- $z = -0.8416$, $\mu = 50 + 5(0.8416) = 54.2$ g
 - $1 - 0.8^8 = 0.832$
- $P(T > 75) = 0.1217 \approx 0.122$
 - $B(8, 0.1217)$: 0.190

(c) $1 - 0.8783^8 = 0.646$

(d) $\frac{75 - 68}{1.6449} = 4.26$ minutes

12. (a) (3, 22.6) ($\bar{y} = 22.57$)

(b) $y = 7.29x + 0.714$

(c) For each extra 1 cm of length, the number of leaves increases by about 7.3

(d) $y \approx 37$ (37.1); $x = 5$ is outside the data range (extrapolation)

(e) Correlation does not imply causation; another factor (e.g. age, light) may affect both.

Need this topic taught, not just practised? Book a trial lesson at 50% off with Mr. Ahmad — online worldwide or in person in Dubai.

WhatsApp +971 55 373 5695 · mathandsciencetutor.com