

IB Maths AI SL – Topic 4: Statistics and Probability

IB AI SL · Total marks: 66 · Answers on the last page

Covers sampling, descriptive statistics, correlation and regression, probability, the binomial and normal distributions, and chi-squared and t-tests from Topic 4 of the IB AI SL syllabus. A GDC is required throughout; give answers to 3 significant figures unless exact. Allow about 90 minutes.

Section A – Data, correlation and probability (GDC allowed)

1. A school has 900 students. The head teacher wants to survey 60 students about the canteen. For each method, state the type of sampling being used.
 - (a) Every 15th student on the alphabetical school register is chosen.
 - (b) The number of students chosen from each year group is proportional to the size of that year group, and the students are chosen at random within each year group.
 - (c) The first 60 students to arrive at school on Monday are chosen. [3]
2. The numbers of books read in a year by seven students were:
12, 15, 15, 18, 21, 24, 30
 - (a) Find the mean and the median of the data.
 - (b) Find the interquartile range.
 - (c) Determine, using the criterion $Q_3 + 1.5 \times \text{IQR}$, whether the value 30 is an outlier. [5]
3. The table shows the number of hours x that six students spent revising and their test score y (percent):
 x : 2, 3, 5, 6, 8, 9
 y : 45, 50, 58, 62, 72, 78
 - (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 revised for 7 hours. Comment on the reliability of this estimate.
 - (d) Explain why the line should not be used to predict the score of a student who revised for 20 hours. [8]
4. A bag contains 5 red and 3 blue counters. Two counters are taken at random without replacement.
 - (a) Find the probability that both counters are red.
 - (b) Find the probability that the counters are different colours.
 - (c) Given that the first counter is blue, find the probability that the second is red. [5]

Section B – Distributions (GDC allowed)

5. In a group of 40 students, 22 study French, 18 study Spanish and 6 study neither language.
 - (a) Find the number of students who study both languages.
 - (b) A student is chosen at random. Find the probability that the student studies French given that they study Spanish.
 - (c) Determine whether the events 'studies French' and 'studies Spanish' are independent. Justify your answer. [5]
6. A factory finds that 12 percent of the light bulbs it produces are faulty. A random sample of 20 bulbs is taken. Let X be the number of faulty bulbs in the sample.
 - (a) State the distribution of X and one assumption required.
 - (b) Find $P(X = 2)$.

- (c) Find $P(X \geq 3)$.
(d) Find the expected value and variance of X . [7]
7. The heights of adult men in a town are normally distributed with mean 172 cm and standard deviation 6.5 cm.
(a) Find the probability that a randomly chosen man is taller than 180 cm.
(b) Find the probability that a man's height is between 165 cm and 175 cm.
(c) Find the height below which the shortest 10 percent of men fall. [6]
8. The times taken to complete a puzzle are normally distributed with mean 40 minutes and standard deviation σ . It is known that 80 percent of people finish within 45 minutes. Find σ . [4]
9. A game costs 3 dollars to play. The amount X dollars won has the probability distribution:
 $P(X = 0) = 0.4$, $P(X = 2) = 0.3$, $P(X = 5) = 0.2$, $P(X = 10) = k$.
(a) Find the value of k .
(b) Find $E(X)$.
(c) Determine whether the game is fair, and find the expected gain or loss for a player per game. [5]

Section C – Hypothesis tests (GDC allowed)

10. A researcher asks 120 students which activity they prefer. The results are:
Boys: Sport 30, Music 20, Art 10. Girls: Sport 20, Music 25, Art 15.
A χ^2 test for independence is carried out at the 5 percent significance level.
(a) State the null and alternative hypotheses.
(b) Write down the number of degrees of freedom.
(c) Show that the expected frequency for boys who prefer Music is 22.5.
(d) Find the χ^2 statistic and the p -value.
(e) State the conclusion of the test in context. [7]
11. A six-sided die is rolled 60 times. The frequencies of the scores 1 to 6 are 8, 12, 9, 11, 13 and 7 respectively. A χ^2 goodness of fit test is used to test, at the 5 percent level, whether the die is fair.
(a) State the expected frequency for each score.
(b) Calculate the χ^2 statistic.
(c) The critical value is 11.07. State the conclusion of the test. [5]
12. Two groups of six students sit the same test. Group A used a new revision app; Group B did not. Group A scores: 62, 70, 75, 68, 80, 72. Group B scores: 58, 65, 60, 70, 66, 63.
Assuming the scores are normally distributed with equal variances, a two-sample t -test is carried out at the 5 percent significance level to decide whether the app improves scores.
(a) State suitable hypotheses in terms of the population means μ_A and μ_B .
(b) Find the mean score of each group.
(c) Find the p -value of the test.
(d) State the conclusion in context. [6]

Answers

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

- Systematic sampling
 - Stratified sampling
 - Convenience sampling
- Mean = $\frac{135}{7} = 19.3$ (3 s.f.); median = 18
 - $Q_1 = 15$, $Q_3 = 24$, IQR = 9
 - Upper fence = $24 + 13.5 = 37.5 > 30$, so 30 is not an outlier
- $r = 0.998$; very strong positive (linear) correlation
 - $y = 4.60x + 35.5$
 - $y = 4.6(7) + 35.53 = 67.7$, so about 68 percent; reliable because it is interpolation and r is close to 1
 - $x = 20$ is outside the range of the data (extrapolation); the relationship may not continue (and the score cannot exceed 100)
- $\frac{5}{8} \times \frac{4}{7} = \frac{5}{14}$
 - $2 \times \frac{5}{8} \times \frac{3}{7} = \frac{15}{28}$
 - $\frac{5}{7}$
- $22 + 18 - 34 = 6$
 - $P(F | S) = \frac{6}{18} = \frac{1}{3}$
 - $P(F)P(S) = 0.55 \times 0.45 = 0.2475$
 $eP(F \cap S) = 0.15$, so not independent
- $X \sim B(20, 0.12)$; bulbs are independent (or constant probability of being faulty)
 - $P(X = 2) = 0.274$
 - $P(X \geq 3) = 1 - P(X \leq 2) = 0.437$
 - $E(X) = 2.4$, $\text{Var}(X) = 20 \times 0.12 \times 0.88 = 2.11$
- $P(X > 180) = 0.109$
 - $P(165 < X < 175) = 0.537$
 - 164 cm (163.7 cm)
- $P(Z < z) = 0.8 \Rightarrow z = 0.8416$; $\frac{45 - 40}{\sigma} = 0.8416 \Rightarrow \sigma = 5.94$ minutes
- $k = 0.1$
 - $E(X) = 0 + 0.6 + 1.0 + 1.0 = 2.6$ dollars
 - Not fair since $E(X) < 3$; expected loss = $3 - 2.6 = 0.40$ dollars per game
- H_0 : preferred activity is independent of gender; H_1 : preferred activity is not independent of gender
 - $(2 - 1)(3 - 1) = 2$
 - $\frac{60 \times 45}{120} = 22.5$
 - $\chi^2_{\text{calc}} = 3.56$, $p = 0.169$
 - $p > 0.05$ (or $3.56 < 5.991$), do not reject H_0 ; there is insufficient evidence that preference depends on gender
- 10 for each score
 - $\chi^2 = \frac{4+4+1+1+9+9}{10} = 2.8$ ($p = 0.731$)

- (c) $2.8 < 11.07$, do not reject H_0 ; no evidence that the die is biased
12. (a) $H_0: \mu_A = \mu_B$; $H_1: \mu_A > \mu_B$ (one-tailed)
(b) $\bar{x}_A = 71.2$, $\bar{x}_B = 63.7$
(c) $t = 2.45$, $p = 0.0173$
(d) $p < 0.05$, reject H_0 ; there is evidence that the app improves mean scores

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