MEGA LECTURE

IB DP MATHEMATICS

AI Topic 4 · Statistics and Probability

MARK SCHEME

4.1 Descriptive Statistics · 4.2 Correlation & Regression
4.3 Probability · 4.4 Probability Distributions

Standard and Higher Level · Syllabus 2025

Fahad H. Ahmad

+92 323 509 4443 | Megalecture.com

Original exam-style questions written for the IB Diploma Programme Mathematics course (first assessment 2025)

Section A — Multiple Choice: Answer Key (15 marks)

1. Answer: **C** — The median is the middle value of a data set when the values are arranged in ascending (or descending) order.

2. Answer: **B** — The mode is the value that occurs most often; here, 7 occurs twice, more than any other value.

3. Answer: **C** — Standard deviation quantifies how spread out the data values are around the mean.

4. Answer: **B** — The median depends only on the position of values (not their size), so it is far less affected by extreme outliers than the mean.

5. Answer: **A** — A value of r close to $+1$ (such as 0.95) indicates a strong positive linear correlation.

6. Answer: **C** — A value of r close to 0 indicates little or no linear correlation between the variables.

7. Answer: **C** — In $y = a+bx$, b is the gradient of the line, representing the change in y for each one-unit increase in x.

8. Answer: **B** — Extrapolation is the (often unreliable) use of a regression model to predict values outside the range of the original data.

9. Answer: **B** — Two events are independent if and only if $P(A\cap B) = P(A)\times P(B)$.

10. Answer: **C** — The addition rule states $P(A\cup B) = P(A)+P(B)-P(A\cap B)$, which avoids double-counting the overlap.

11. Answer: **A** — Conditional probability is defined as $P(A|B) = P(A\cap B)\div P(B)$.

12. Answer: **B** — Mutually exclusive events cannot both happen at the same time, so $P(A\cap B) = 0$.

13. Answer: **B** — A binomial distribution models a fixed number of independent trials, each with exactly two outcomes ('success'/'failure') and a constant probability of success.

14. Answer: **A** — For a binomially distributed random variable $X\sim B(n,p)$, the mean is $E(X) = np$.

15. Answer: **B** — For a binomially distributed random variable $X\sim B(n,p)$, the variance is $Var(X) = np(1-p)$.

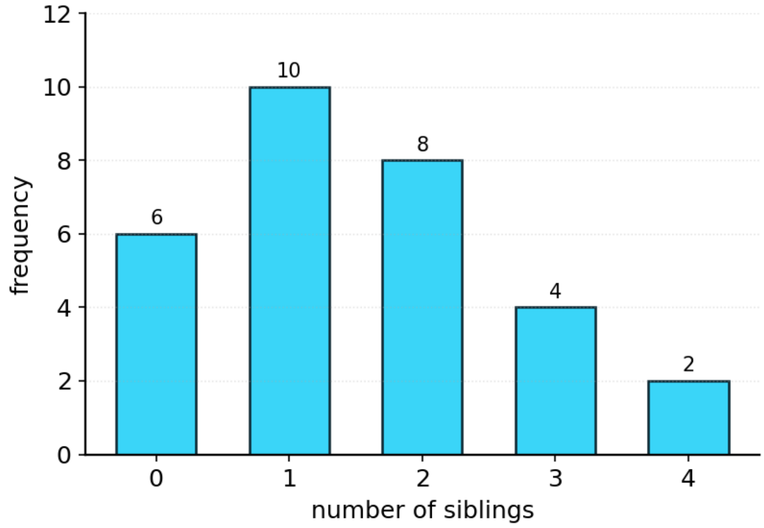
Section B — Structured Questions: Worked Solutions (65 marks)

Mark-scheme notation: M = method mark, A = accuracy mark (dependent on the preceding M mark), R = reasoning mark.

4.1 Descriptive Statistics

Question 1 [7 marks]

A survey of 30 students records the number of siblings each student has. The bar chart shows the frequency distribution (given data).



Frequency distribution of number of siblings, n=30 (given data).

(a) Using the bar chart, write down the total number of students surveyed, n. **[1]**

n = **30** students. **A1**

(b) Calculate the mean number of siblings, giving your answer to 3 significant figures. **[3]**

Mean = (0×6+1×10+2×8+3×4+4×2)÷30 = 46÷30 = **1.53**. **M1 M1 A1**

(c) Write down the median number of siblings, and calculate the standard deviation, giving your answer to 3 significant figures. **[3]**

Median = **1** (the 15th and 16th values, in order, are both 1). Standard deviation = **1.15** (using technology). **A1 M1 A1**

Question 2 [6 marks]

The exam scores (out of 100) of 8 students are: 62, 75, 58, 91, 68, 73, 80, 55.

(a) Calculate the range of the scores. **[1]**

Range = 91−55 = **36**. **A1**

(b) Arrange the data in order, and hence find the lower quartile, Q_1, and upper quartile, Q_3. **[3]**

Ordered data: 55, 58, 62, 68, 73, 75, 80, 91. Q_1 (median of lower half 55,58,62,68) = **60.0**; Q_3 (median of upper half 73,75,80,91) = **77.5**. **M1 A1 A1**

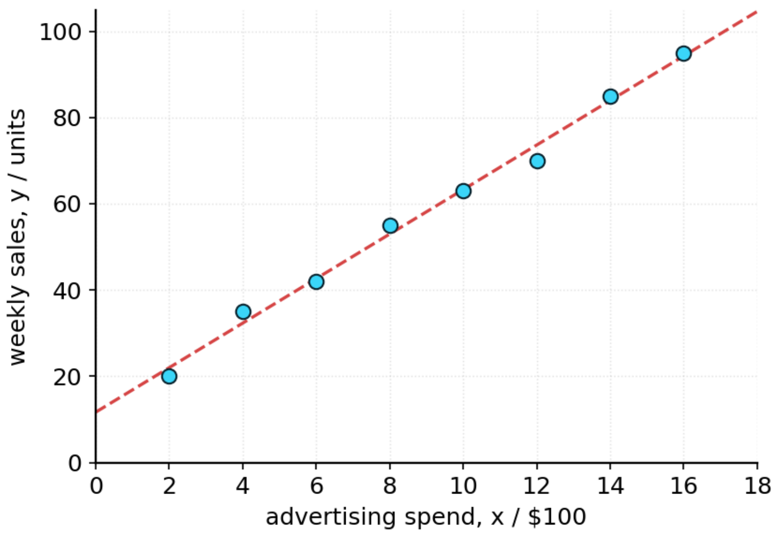
(c) Using the interquartile range and the 1.5×IQR rule, determine whether the data set contains any outliers. **[2]**

IQR = 77.5−60.0 = 17.5. Fences: Q_1−1.5·IQR = 33.75, Q_3+1.5·IQR = 103.75. Since all scores lie within [33.75, 103.75], **there are no outliers**. **M1 A1**

4.2 Correlation & Regression

Question 3 [7 marks]

The table shows the weekly advertising spend, x (in $100s), and the weekly sales, y (units), for 8 stores. The scatter diagram shows this data, with a line of best fit (given data).



Advertising spend vs. weekly sales, n=8 stores, with regression line (given data).

(a) Using technology, write down the value of Pearson's product-moment correlation coefficient, r, and comment on the strength and direction of the correlation. **[2]**

r = **0.997**, indicating a very strong positive linear correlation between advertising spend and weekly sales. **A1 A1**

(b) Using technology, write down the equation of the regression line of y on x, giving values to 3 significant figures. **[2]**

y = **5.16x+11.7** (using technology). **A1 A1**

(c) Use the regression equation to predict the weekly sales for an advertising spend of $1300 (x=13), giving your answer to the nearest unit. **[3]**

y = 5.161(13)+11.679 = **79 units** (to the nearest unit). **M1 M1 A1**

Question 4 [6 marks]

The regression line $y = 3.2x+12.5$ models the relationship between the number of hours studied, x, and the exam score, y (out of 100), for a group of students.

(a) Interpret, in context, what the gradient of this regression line represents. **[2]**

The gradient (3.2) means that **for each additional hour studied, the exam score is predicted to increase by 3.2 marks**. **R1 R1**

(b) Interpret, in context, what the y-intercept of this regression line represents. **[2]**

The y-intercept (12.5) represents **the predicted exam score for a student who studies for 0 hours**. **R1 R1**

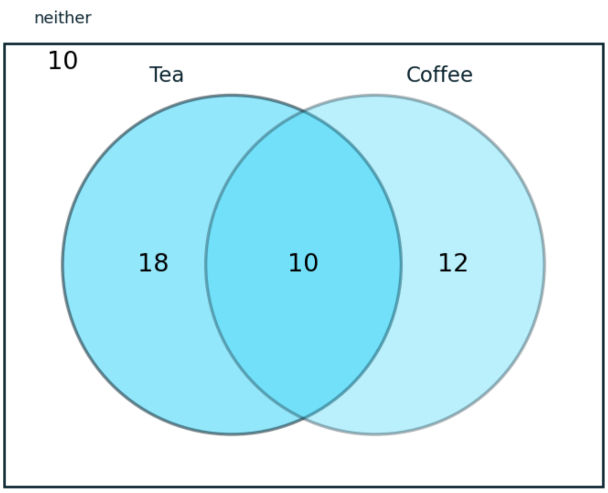
(c) Use the regression equation to predict the exam score for a student who studies for 10 hours. **[2]**

$y = 3.2(10)+12.5 =$ **44.5**. **M1 A1**

4.3 Probability

Question 5 [7 marks]

In a survey of 50 people, 28 like tea, 22 like coffee, and 10 like both tea and coffee. The Venn diagram shows this information (given data).



Venn diagram, tea/coffee survey of 50 people (given data).

(a) Using the Venn diagram, write down the number of people who like tea only (and not coffee). **[2]**

Tea only = $28-10$ = **18** people. **M1 A1**

(b) Write down the number of people who like neither tea nor coffee. **[2]**

Neither = $50-(18+10+12)$ = **10** people. **M1 A1**

(c) A person who likes coffee is chosen at random. Calculate the probability that this person also likes tea, i.e. find P(tea|coffee). **[3]**

$P(\text{tea}|\text{coffee}) = P(\text{tea}\cap\text{coffee}) \div P(\text{coffee}) = 10 \div 22 =$ **0.455** (3 s.f.). **M1 M1 A1**

Question 6 [6 marks]

A bag contains 5 red balls and 3 blue balls. Two balls are drawn at random from the bag, one after another, without replacement.

(a) Write down the probability that the first ball drawn is red. **[1]**

P(first red) = 5÷8 = **0.625**. **A1**

(b) Calculate the probability that both balls drawn are red. **[3]**

P(both red) = 5÷8×4÷7 = **0.357** (3 s.f.). **M1 M1 A1**

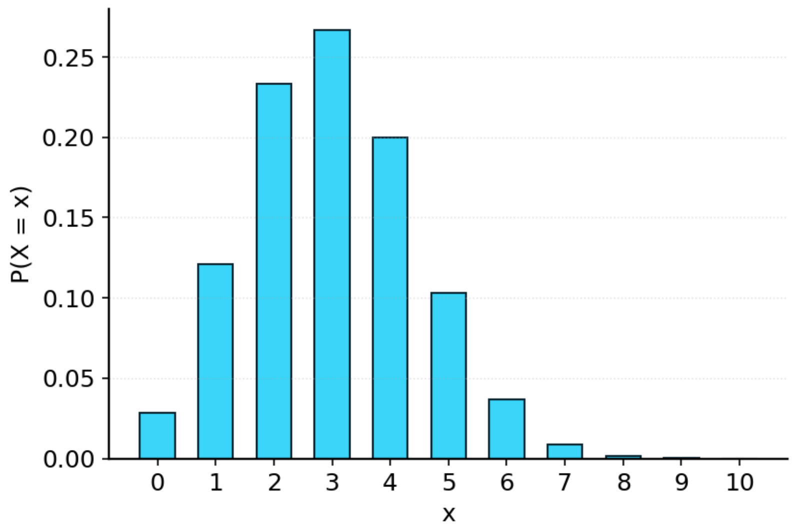
(c) Hence, calculate the probability that at least one of the two balls drawn is blue. **[2]**

P(at least one blue) = 1−P(both red) = 1−0.357 = **0.643** (3 s.f.). **M1 A1**

4.4 Probability Distributions

Question 7 [7 marks]

A random variable X follows a binomial distribution X~B(10, 0.3). The graph shows the probability distribution of X (computed).



Binomial distribution, X~B(10, 0.3) (computed).

(a) Using the graph (or the parameters of the distribution), write down the number of trials, n, and the probability of success, p. **[2]**

n = **10**; p = **0.3**. **A1 A1**

(b) Using technology, calculate P(X=3), giving your answer to 3 significant figures. **[2]**

P(X=3) = **0.267** (using technology). **M1 A1**

(c) Calculate the mean and variance of X. **[3]**

Mean = np = 10×0.3 = **3.0**. Variance = np(1−p) = 10×0.3×0.7 = **2.1**. **M1 A1 A1**

Question 8 [6 marks]

A multiple-choice quiz has 5 questions, each with 4 possible answers, only one of which is correct. A student guesses randomly on every question, so the number of correct answers, X, follows a binomial distribution $X \sim B(5, 0.25)$.

(a) Using technology, calculate the probability that the student answers exactly 2 questions correctly, giving your answer to 3 significant figures. **[2]**

$P(X=2)$ = **0.264** (using technology). **M1 A1**

(b) Calculate the probability that the student answers at least 1 question correctly, giving your answer to 3 significant figures. **[2]**

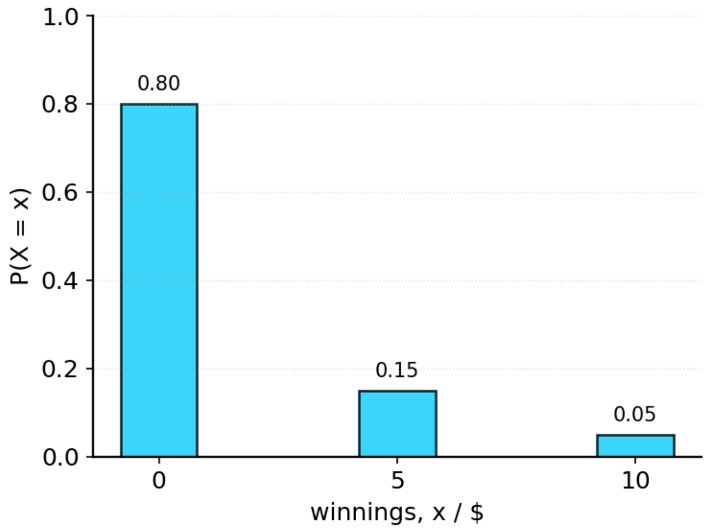
$P(X\geq 1) = 1-P(X=0)$ = **0.763** (3 s.f.). **M1 A1**

(c) Calculate the expected number of questions the student answers correctly. **[2]**

$E(X) = np = 5\times 0.25$ = **1.25**. **M1 A1**

Question 9 [7 marks]

A game costs \$2 to play. A player wins \$10 with probability 0.05, wins \$5 with probability 0.15, and wins \$0 with probability 0.80. The graph shows the probability distribution of the winnings, X (given data).



Probability distribution of game winnings, X (given data).

(a) Verify that the probabilities in the distribution sum to 1. **[1]**

$0.05+0.15+0.80$ = **1.00**, confirming this is a valid probability distribution. **A1**

(b) Calculate the expected value (mean) of the winnings, E(X). **[3]**

$E(X) = 10\times 0.05+5\times 0.15+0\times 0.80 = 0.5+0.75+0$ = **\$1.25**. **M1 M1 A1**

(c) Given that it costs \$2 to play the game, calculate the expected profit (or loss) per game for the player. **[3]**

Expected profit = $E(X) - \text{cost} = 1.25-2$ = **−\$0.75**, i.e. an expected **loss of \$0.75 per game**. **M1 A1 R1**

4.3 Probability

Question 10 [6 marks]

A survey of 60 students records their gender and their preferred activity (sport or music). Of the 30 male students, 18 prefer sport and 12 prefer music. Of the 30 female students, 10 prefer sport and 20 prefer music.

(a) Construct a two-way (contingency) table for this data, and write down the total number of students who prefer sport, and the total number who prefer music. **[2]**

Total preferring sport = 18+10 = **28**; total preferring music = 12+20 = **32**. **M1 A1**

(b) A male student is chosen at random. Calculate the probability that this student prefers sport, i.e. find P(sport|male). **[2]**

P(sport|male) = 18÷30 = **0.60**. **M1 A1**

(c) A student who prefers music is chosen at random. Calculate the probability that this student is female, i.e. find P(female|music). **[2]**

P(female|music) = 20÷32 = **0.625** (3 s.f.). **M1 A1**