

IB DP MATHEMATICS

AI Topic 4 · Statistics and Probability

TOPICAL WORKSHEET

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

Section A — Multiple Choice (15 marks)

Circle the letter of the best answer to each question. Each question is worth 1 mark.

1. The median of a data set is best described as [1]
A the most frequently occurring value
B the sum of all values divided by the number of values
C the middle value when the data is arranged in order
D the difference between the largest and smallest values
2. For the data set 3, 7, 7, 9, 12, the mode is [1]
A 3
B 7
C 9
D 12
3. The standard deviation of a data set measures [1]
A the middle value of the data
B the most common value in the data
C the spread (dispersion) of the data around the mean
D the total number of data values
4. A data set contains an extreme outlier. The measure of central tendency LEAST affected by this outlier is the [1]
A mean
B median
C range
D sum
5. A correlation coefficient of $r=0.95$ between two variables indicates [1]
A a strong positive linear correlation
B a strong negative linear correlation
C no linear correlation
D a weak positive correlation
6. A correlation coefficient of $r=0.03$ between two variables indicates [1]
A a strong positive correlation
B a strong negative correlation
C little or no linear correlation
D a perfect correlation

- 7.** In the regression line $y = a + bx$, the coefficient b represents [1]
 A the y-intercept
 B the predicted value of y when $x=0$
 C the rate of change of y per unit increase in x (the gradient)
 D the correlation coefficient
- 8.** Using a regression line to predict a y -value for an x -value far outside the range of the original data is called [1]
 A interpolation
 B extrapolation
 C correlation
 D regression
- 9.** Two events A and B are independent if [1]
 A $P(A \cap B) = P(A) + P(B)$
 B $P(A \cap B) = P(A) \times P(B)$
 C $P(A \cap B) = 0$
 D $P(A|B) = P(B)$
- 10.** For any two events A and B , $P(A \cup B)$ is given by [1]
 A $P(A) + P(B)$
 B $P(A) \times P(B)$
 C $P(A) + P(B) - P(A \cap B)$
 D $P(A) - P(B)$
- 11.** The conditional probability $P(A|B)$ is defined as [1]
 A $P(A \cap B) \div P(B)$
 B $P(A) \div P(B)$
 C $P(A \cap B) \times P(B)$
 D $P(B) \div P(A)$
- 12.** Two events are described as mutually exclusive if [1]
 A they always occur together
 B they cannot occur at the same time, i.e. $P(A \cap B) = 0$
 C they are independent of each other
 D $P(A) = P(B)$
- 13.** A binomial distribution requires [1]
 A a continuous random variable with infinitely many outcomes
 B a fixed number of independent trials, each with two outcomes and a constant probability of success
 C a variable number of trials with changing probability of success
 D at least three possible outcomes per trial

14. For a random variable $X \sim B(n, p)$, the expected value (mean) $E(X)$ is given by **[1]**

- A np
- B $np(1-p)$
- C $n+p$
- D $\sqrt{np(1-p)}$

15. For a random variable $X \sim B(n, p)$, the variance $\text{Var}(X)$ is given by **[1]**

- A np
- B $np(1-p)$
- C $n(1-p)$
- D $p(1-p)$

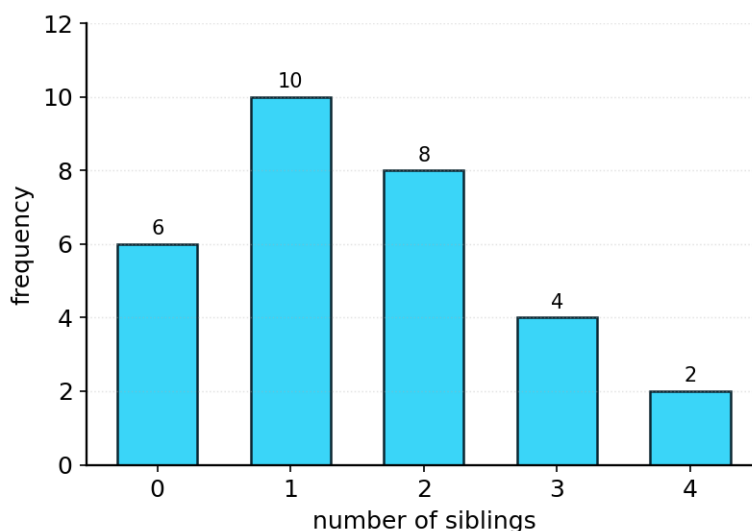
Section B — Structured Questions (65 marks)

Answer all questions in the space provided. Working must be shown for full marks.

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]

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

[3]

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

[3]

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]

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

[3]

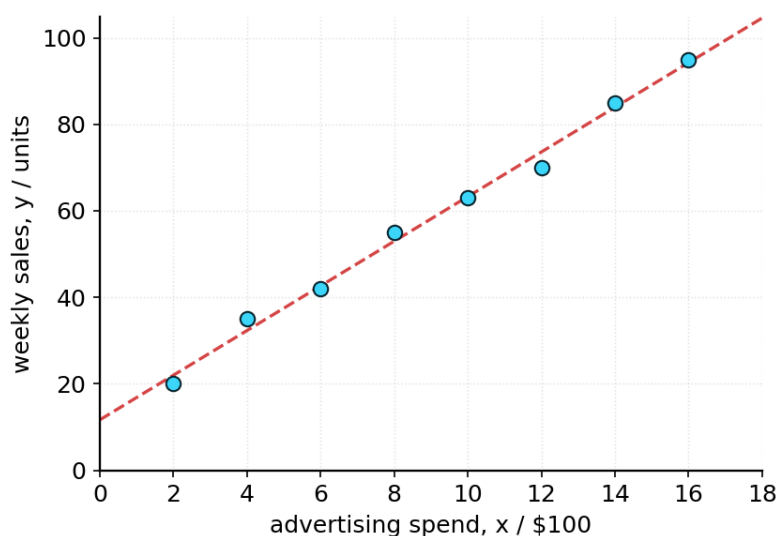
(c) Using the interquartile range and the $1.5 \times \text{IQR}$ rule, determine whether the data set contains any outliers.

[2]

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]

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

[2]

(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]

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]

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

[2]

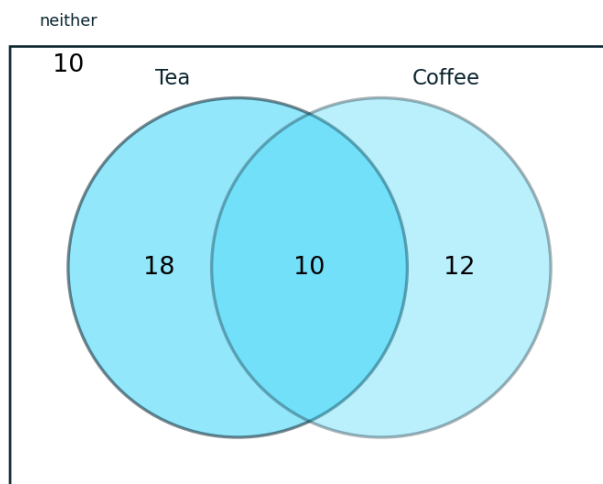
(c) Use the regression equation to predict the exam score for a student who studies for 10 hours.

[2]

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]**

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

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

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]**

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

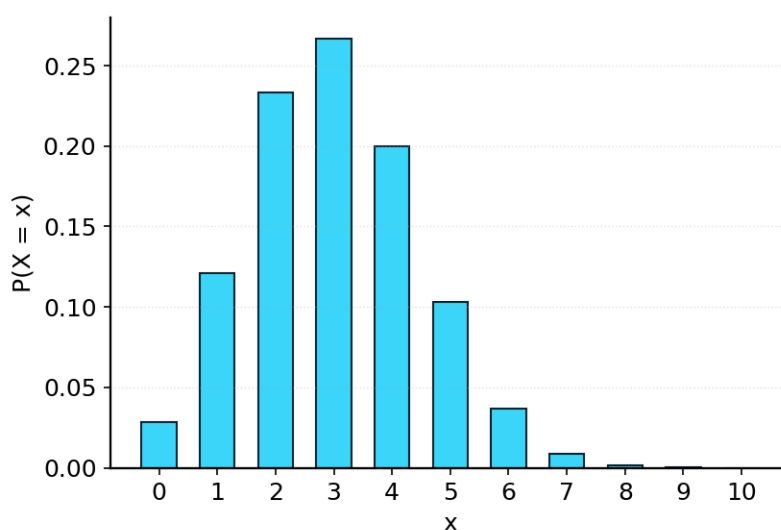
(c) Hence, calculate the probability that at least one of the two balls drawn is blue.

[2]

4.4 Probability Distributions

Question 7 [7 marks]

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



Binomial distribution, $X \sim B(10, 0.3)$ (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]

(b) Using technology, calculate $P(X=3)$, giving your answer to 3 significant figures.

[2]

(c) Calculate the mean and variance of X .

[3]

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 $X \sim B(5, 0.25)$.

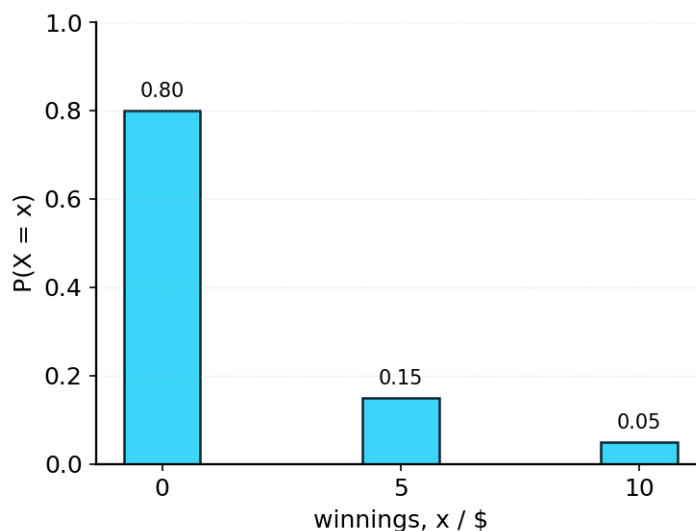
- (a)** Using technology, calculate the probability that the student answers exactly 2 questions correctly, giving your answer to 3 significant figures. [2]

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

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

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]

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

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

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]

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

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