eClassroom

INTERNATIONAL A LEVEL MATHEMATICS • STATISTICS 1 (WST01)

Statistics 1 Practice Paper 1

Worked Solutions

AIMING FOR B

Written in the style of the Pearson Edexcel International Advanced Level paper (WST01/01) • 75 marks

Every question solved in full on the paper itself, the way your teacher would write it out

How to read these solutions. Each page is the real question page, with a full answer written on the answer lines.

Red pen

M1 A1

boxed

the working a full-marks answer shows, with a short note on why each step is taken

the mark earned on that line: M = method, A = accuracy, B = independent mark

the final answer to each part

Sit the paper first. These solutions are most useful after you have tried every question yourself under timed conditions. Mark your paper with the mark scheme, then use this booklet to see exactly how each lost mark could have been earned.

Q	Topic	Marks	Page
1	A discrete random variable	10	2
2	Stem-and-leaf diagram and box plot	10	4
3	Correlation and regression	11	6
4	A Venn diagram and independence	9	8
5	Histogram and grouped data	12	9
6	The normal distribution	11	11
7	Tree diagram without replacement	12	12

Question paper & mark scheme | Exam questions by topic | More practice papers

1. The discrete random variable X has probability function

$$\mathrm{P}(X = x) = \begin{cases} kx & x = 1, 2, 3 \\ k(7 - x) & x = 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{1}{12}$

(2)

(b) Find F(2.5)

(2)

(c) Write down the value of $\mathrm{E}(X)$, giving a reason for your answer.

(1)

(d) Show that $\mathrm{Var}(X) = \frac{23}{12}$

(3)

(e) Find $\mathrm{Var}(5 - 3X)$

(2)

(a) The probabilities must add up to 1: $k + 2k + 3k + 3k + 2k + k = 1$ M1

$12k = 1 \Rightarrow k = 1/12$ ✓ A1*

(b) X only takes whole numbers, so $F(2.5) = P(X \leqslant 2) = P(X = 1) + P(X = 2)$ M1

$= 1/12 + 2/12 = 3/12$ $F(2.5) = \frac{1}{4}$ A1

(c) $P(X = x)$ is symmetrical about 3.5 (1, 2, 3, 3, 2, 1) $E(X) = 3.5$ B1

(d) $E(X^2) = (1\times1 + 4\times2 + 9\times3 + 16\times3 + 25\times2 + 36\times1) \div 12$ M1

$= (1 + 8 + 27 + 48 + 50 + 36) \div 12 = 170/12 = 85/6$ A1

$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 85/6 - 3.5^2 = 170/12 - 147/12 = 23/12$ ✓ A1*

(e) $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$: the 5 does not change the spread, and −3 is squared M1

$\mathrm{Var}(5 - 3X) = 9 \times 23/12 = 207/12$ $\mathrm{Var}(5 - 3X) = 69/4 = 17.25$ A1

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 1 continued

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

(Total for Question 1 is 10 marks)

★★★★★

Join students already acing their exams.

Book your FREE lesson at **www.eclassroom.org.uk**

2.

Key: 2 | 3 means 23 minutes

1	2 5 7 8	(4)
2	0 1 3 3 6 8 9	(7)
3	1 2 4 4 5 7 8	(7)
4	1 3 7 8 9	(5)
5	3 7 8	(3)
6		(0)
7		(0)
8	6	(1)

Figure 1

The stem and leaf diagram in Figure 1 shows the waiting times, in minutes, of 27 patients at a clinic.

(a) Find the median and the quartiles of these waiting times.

(3)

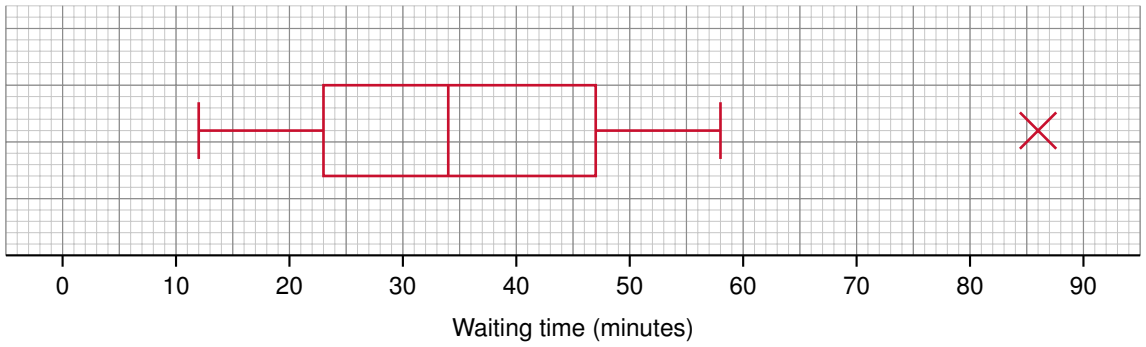
An outlier is defined as a value that is greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$

(b) Show that 86 is an outlier, and that there are no other outliers.

(2)

(c) On the grid below, draw a box plot to represent these waiting times.

(3)



(d) Describe the skewness of the waiting times, giving a reason for your answer.

(2)

(a) $n = 27$. Median: $(27 + 1)/2 = 14$th value $Q_2 = 34$ B1

$27/4 = 6.75$, not a whole number, so round up: 7th value $Q_1 = 23$ B1

$3 \times 27/4 = 20.25$, so round up: 21st value $Q_3 = 47$ B1

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 2 continued

(b) $IQR = 47 - 23 = 24$. Upper limit: $47 + 1.5 \times 24 = 83$. Lower limit: $23 - 36 = -13$ M1

$86 > 83$, so 86 is an outlier ✓. The next largest is $58 < 83$, and $12 > -13$: no others. A1

(c) Drawn on the grid above. Box from $Q_1 = 23$ to $Q_3 = 47$, median at 34 B1

Whiskers to 12 and to 58: the largest value that is not an outlier, not 86 B1

The outlier 86 is marked on its own with a cross B1

(d) $Q_3 - Q_2 = 13$ and $Q_2 - Q_1 = 11$, so $13 > 11$ B1

positive skew B1

(Total for Question 2 is 10 marks)

DO NOT WRITE IN THIS AREA

3. On each of 10 days in spring, a café records the maximum temperature, t °C, and the number of hot drinks it sells, d

The results are summarised below.

$$\sum t = 112 \qquad \sum d = 770 \qquad \sum t^2 = 1470 \qquad \sum d^2 = 62446 \qquad \sum td = 7829$$

$$S_{tt} = 215.6 \qquad S_{dd} = 3156$$

(a) Show that $S_{td} = -795$

(2)

(b) Find the product moment correlation coefficient between t and d.

(2)

(c) Interpret your value of the correlation coefficient in the context of the question.

(1)

(d) Find the equation of the regression line of d on t, in the form $d = a + bt$, giving the values of a and b to 3 significant figures.

(3)

(e) Use your regression line to estimate the number of hot drinks sold on a day when the maximum temperature is 12 °C.

(1)

The lowest maximum temperature on these 10 days was 4 °C and the highest was 19 °C.

(f) Explain why the regression line should not be used to estimate the number of hot drinks sold on a day when the maximum temperature is 30 °C.

(1)

(g) The café owner says that the regression line will give the exact number of hot drinks sold on any day. Give a reason why a statistical model like this can only be an approximation.

(1)

(a) $S_{td} = \sum td - (\sum t)(\sum d)/n = 7829 - (112 \times 770)/10$ M1

$= 7829 - 8624 = -795$ ✓ A1*

(b) $r = S_{td} / \sqrt{S_{tt}S_{dd}} = -795 / \sqrt{215.6 \times 3156}$ M1

$= -795 / 824.88...$ $r = -0.964$ A1

(c) Strong negative correlation: on warmer days, fewer hot drinks are sold. B1

(d) $b = S_{td} / S_{tt} = -795 / 215.6 = -3.6873...$ M1

$\bar{t} = 11.2$, $\bar{d} = 77$: $a = \bar{d} - b\bar{t} = 77 + 3.6873... \times 11.2 = 118.29...$ M1

$d = 118 - 3.69t$ A1

(e) $d = 118.29... - 3.6873... \times 12 = 74.05...$ about 74 hot drinks B1

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 3 continued

(f) 30 °C is outside the range 4 °C to 19 °C, so it would be extrapolation. B1

(g) A model simplifies the real world: rain or the day of the week also affect sales. B1

(Total for Question 3 is 11 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

★★★★★
Join students already acing their exams.
Book your FREE lesson at www.eclassroom.org.uk

4.

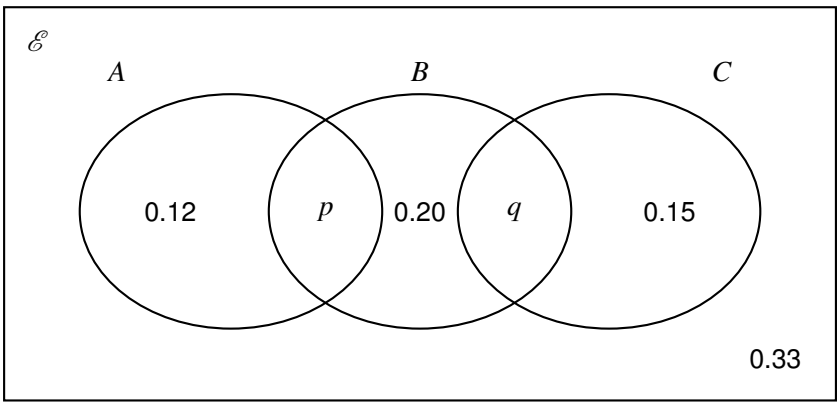


Figure 2

The Venn diagram in Figure 2 shows the probabilities associated with three events, A, B and C

The events A and C are mutually exclusive, and p and q are probabilities.

Given that the events A and B are independent,

(a) find the value of p and the value of q

(4)

(b) Find P $\left(B \mid A \cup C\right)$

(3)

(c) Determine whether or not the events B and C are independent. Show your working clearly.

(2)

(a) Total probability 1: $0.12 + p + 0.20 + q + 0.15 + 0.33 = 1 \Rightarrow p + q = 0.2$ M1

So $P(B) = p + 0.20 + q = 0.4$, and $P(A) = 0.12 + p$

Independent: $P(A \cap B) = P(A) \times P(B) \Rightarrow p = (0.12 + p) \times 0.4$ M1

$p = 0.048 + 0.4p \Rightarrow 0.6p = 0.048$ $\boxed{p = 0.08}$ A1

$q = 0.2 - 0.08$ $\boxed{q = 0.12}$ A1

(b) $P(A \cup C) = 0.12 + 0.08 + 0.12 + 0.15 = 0.47$ M1

$P(B \cap (A \cup C)) = 0.08 + 0.12 = 0.2$ (the parts of B inside A or C) M1

$P(B \mid A \cup C) = 0.2 / 0.47$ $\boxed{20/47 = 0.426}$ A1

(c) $P(B) \times P(C) = 0.4 \times (0.12 + 0.15) = 0.4 \times 0.27 = 0.108$ M1

$P(B \cap C) = 0.12 \neq 0.108$, so B and C are $\boxed{\text{not independent}}$ A1

(Total for Question 4 is 9 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

★★★★★
Join students already acing their exams.
Book your FREE lesson at www.eclassroom.org.uk

5.

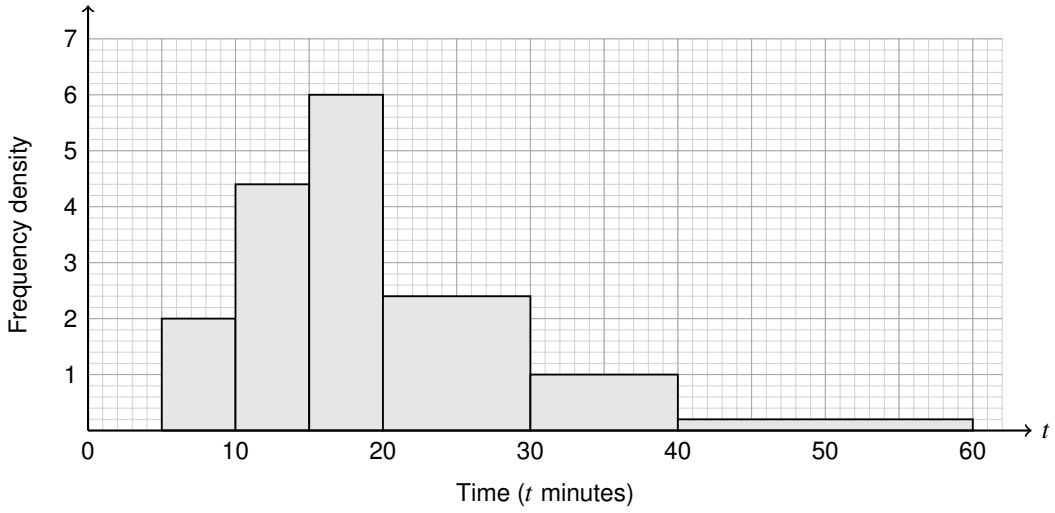


Figure 3

The histogram in Figure 3 shows the times, t minutes, taken by 100 people to complete a puzzle.

(a) Find the number of people who took less than 15 minutes.

(2)

(b) Estimate the number of people who took between 12 minutes and 25 minutes.

(2)

(c) Use linear interpolation to estimate the median time.

(2)

(d) Using the midpoints of the classes, find estimates of the mean and the standard deviation of the times.

(3)

(e) Describe, giving a reason, the skewness of the times.

(2)

(f) Give a reason why a normal distribution would not be a suitable model for these times.

(1)

(a) Frequency = area = class width × frequency density

$5 \times 2 + 5 \times 4.4 = 10 + 22$ 32 people M1 A1

(b) 12 to 15: $3 \times 4.4 = 13.2$; 15 to 20: $5 \times 6 = 30$; 20 to 25: $5 \times 2.4 = 12$ M1

$13.2 + 30 + 12 = 55.2$ about 55 people A1

(c) Median = 50th value. 32 people took less than 15, and 30 are in 15 to 20

Median $= 15 + (50 - 32)/30 \times 5 = 15 + 3$ 18 minutes M1 A1

(d) Frequencies 10, 22, 30, 24, 10, 4; midpoints 7.5, 12.5, 17.5, 25, 35, 50

$\sum ft = 2025 \Rightarrow$ mean $= 2025/100 = 20.25$ M1

DO NOT WRITE IN THIS AREA

★★★★★ Join students already acing their exams.
Book your FREE lesson at www.eclassroom.org.uk

Question 5 continued

$\sum ft^2 = 50437.5 \Rightarrow \sigma = \sqrt{50437.5/100 - 20.25^2} = \sqrt{94.3125}$ M1

mean = 20.25 minutes s.d. = 9.71 minutes A1

(e) mean (20.25) > median (18) positive skew B1 B1

(f) The times are skewed, but a normal distribution is symmetrical. B1

DO NOT WRITE IN THIS AREA

(Total for Question 5 is 12 marks)

★★★★★
Join students already acing their exams.
Book your FREE lesson at www.eclassroom.org.uk

6. The mass of flour in a bag, X grams, is modelled by a normal distribution with mean 1010 grams and standard deviation 8 grams.

(a) Find the probability that a randomly chosen bag contains less than 1000 grams of flour.

(2)

(b) Write down the probability that a randomly chosen bag contains more than 1020 grams of flour.

(1)

(c) Find the value of w such that $P(X > w) = 0.05$

(3)

Three bags are chosen at random.

(d) Find the probability that exactly two of the three bags contain less than 1000 grams of flour.

(2)

The machine that fills the bags is adjusted so that the mean changes but the standard deviation stays at 8 grams. After the adjustment, 1% of bags contain less than 1000 grams of flour.

(e) Show that the new mean is 1019 grams, to the nearest gram.

(3)

(a) $P(X < 1000) = P(Z < (1000 - 1010)/8) = P(Z < -1.25)$ M1

$= 1 - P(Z < 1.25) = 1 - 0.8944$ 0.1056 A1

(b) 1020 is 10 above the mean, as 1000 is 10 below it (symmetry) 0.1056 B1

(c) $P(X > w) = 0.05 \Rightarrow P(Z < z) = 0.95 \Rightarrow z = 1.6449$ (from the tables) B1

$(w - 1010)/8 = 1.6449$ M1

$w = 1010 + 13.159... = 1023.16...$ w = 1023 grams A1

(d) L = less than 1000 (0.1056), H = not (0.8944). Orders LLH, LHL, HLL: 3 of them M1

$3 \times 0.1056^2 \times 0.8944 = 0.02992...$ 0.0299 A1

(e) $P(X < 1000) = 0.01$: 1000 is below the mean, so $z = -2.3263$ B1

$(1000 - \mu)/8 = -2.3263$ M1

$\mu = 1000 + 8 \times 2.3263 = 1018.61... = 1019$ grams to the nearest gram ✓ A1*

(Total for Question 6 is 11 marks)

DO NOT WRITE IN THIS AREA

7. A bag contains 6 red counters and 4 blue counters. Two counters are taken from the bag at random, one after the other, without replacement.

(a) Complete the tree diagram in Figure 4.

(2)

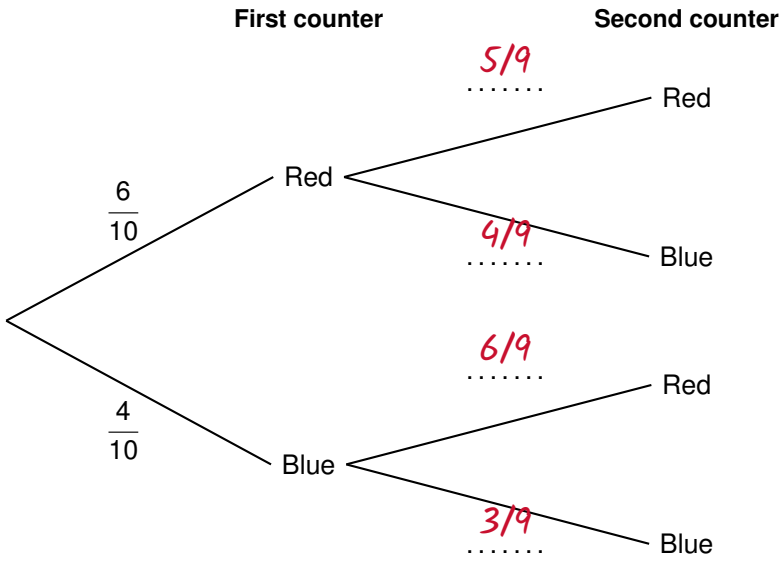


Figure 4

(b) Find the probability that the two counters are the same colour.

(2)

(c) Find the probability that at least one of the counters is blue.

(2)

(d) Given that at least one of the counters is blue, find the probability that the first counter taken is red.

(3)

The counters are put back in the bag, and then k more blue counters are added. Two counters are again taken at random without replacement. The probability that both counters are red is now $\frac{1}{7}$

(e) Show that $k^2 + 19k - 120 = 0$ and hence find the value of k.

(3)

(a) After a red, 9 counters are left: 5 red and 4 blue. After a blue: 6 red and 3 blue.

On the tree above: 5/9 and 4/9 after a red, 6/9 and 3/9 after a blue B1 B1

(b) P(same) = P(RR) + P(BB) = 6/10 × 5/9 + 4/10 × 3/9 = 30/90 + 12/90 M1

= 42/90 7/15 A1

(c) P(at least one blue) = 1 − P(RR) = 1 − 30/90 = 60/90 2/3 M1 A1

(d) Red first and at least one blue means RB: P(RB) = 6/10 × 4/9 = 24/90 = 4/15 M1

DO NOT WRITE IN THIS AREA

Question 7 continued

P(first red | at least one blue) = (4/15) ÷ (2/3) = 12/30 M1

2/5 A1

(e) Now 6 red and 4 + k blue: 10 + k counters. P(RR) = 6/(10 + k) × 5/(9 + k) = 1/7 M1

$(10 + k)(9 + k) = 210 \Rightarrow k^2 + 19k + 90 = 210 \Rightarrow k^2 + 19k - 120 = 0$ ✓ A1*

$(k - 5)(k + 24) = 0$; $k = -24$ is impossible $k = 5$ A1

DO NOT WRITE IN THIS AREA

(Total for Question 7 is 12 marks)

TOTAL FOR PAPER IS 75 MARKS

END

★★★★★
Join students already acing their exams.
Book your FREE lesson at www.eclassroom.org.uk