eClassroom

INTERNATIONAL A LEVEL MATHEMATICS • STATISTICS 1 (WST01)

Statistics 1 Practice Paper 3

Worked Solutions

AIMING FOR A/A*

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 distribution from its cumulative function	12	2
2	Grouped ages with missing frequencies	13	3
3	Independence and a quadratic	12	4
4	Regression with a negative coding	13	6
5	A normal model and rounded data	12	8
6	A back-to-back stem and leaf diagram	13	9

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

★★★★★

Join students already acing their exams.

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

1. The discrete random variable X takes the values $1, 2, 3, \ldots, n$ and has cumulative distribution function

$$F(x) = \frac{x(x+1)}{n(n+1)} \qquad x = 1, 2, 3, \ldots, n$$

(a) Show that $P(X = x) = \frac{2x}{n(n+1)}$ for $x = 1, 2, \ldots, n$

(3)

Given that $E(X) = 7$

(b) find the value of n

(4)

(c) find $\text{Var}(X)$

(3)

(d) find $P(X > E(X))$

(2)

You may use $\sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}$ and $\sum_{x=1}^{n} x^3 = \frac{n^2(n+1)^2}{4}$

(a) F is cumulative, so P(X = x) = F(x) − F(x − 1) M1

= [x(x + 1) − (x − 1)x] ÷ n(n + 1) = (x² + x − x² + x) ÷ n(n + 1) M1

= 2x ÷ n(n + 1) ✓ A1*

(b) E(X) = ∑ x × 2x/(n(n + 1)) = 2/(n(n + 1)) × ∑x² M1

= 2/(n(n + 1)) × n(n + 1)(2n + 1)/6 (using the given formula) M1

= (2n + 1)/3 A1

(2n + 1)/3 = 7 ⇒ 2n + 1 = 21 n = 10 A1

(c) E(X²) = 2/(n(n + 1)) × ∑x³ = 2/(n(n + 1)) × n²(n + 1)²/4 = n(n + 1)/2 M1

n = 10: E(X²) = 110/2 = 55 A1

Var(X) = 55 − 7² Var(X) = 6 A1

(d) X > 7 means X ⩾ 8, so P(X > 7) = 1 − F(7) = 1 − (7 × 8)/(10 × 11) = 1 − 56/110 M1

= 54/110 27/55 A1

(Total for Question 1 is 12 marks)

DO NOT WRITE IN THIS AREA

2. The ages, in completed years, of the 80 members of a walking club are summarised below.

Age (years)	10–19	20–29	30–39	40–49	50–69
Frequency	12	a	22	b	8

An estimate of the mean age, using the midpoints of the classes, is 32.5 years.

(a) Find the value of a and the value of b

(4)

(b) Use linear interpolation to estimate the median age.

(3)

In a histogram of these data, the bar for the class 20–29 is 5.4 cm high.

(c) Find the height of the bar for the class 50–69

(2)

The midpoints, x, of the classes are coded using $y = \dfrac{x - 35}{5}$, and $\sum fy^2 = 544$

(d) Find an estimate of the standard deviation of the ages.

(4)

(a) Ages in completed years: 10–19 means $10 \leqslant$ age < 20, so the midpoint is 15 (not 14.5)

Midpoints 15, 25, 35, 45, 60 B1

Total: $12 + a + 22 + b + 8 = 80 \Rightarrow a + b = 38$ M1

Mean: $15\times12 + 25a + 35\times22 + 45b + 60\times8 = 32.5 \times 80 = 2600 \Rightarrow 25a + 45b = 1170$ M1

$25(a + b) = 950$, so $20b = 220$ $\boxed{a = 27}$ $\boxed{b = 11}$ A1

(b) Cumulative: 12, 39, 61, … The 40th value is in 30–39, boundaries 30 and 40 B1

median $= 30 + (40 - 39)/22 \times 10 = 30 + 0.45\ldots$ M1

$\boxed{\text{median} = 30.5 \text{ years}}$ A1

(c) 20–29: f.d. $27 \div 10 = 2.7$ is 5.4 cm, so 2 cm per unit. 50–69: f.d. $8 \div 20 = 0.4$ M1

height $= 0.4 \times 2$ $\boxed{0.8 \text{ cm}}$ (the class is 20 years wide) A1

(d) $\bar{y} = (\bar{x} - 35)/5 = (32.5 - 35)/5 = -0.5$ M1

$\text{Var}(y) = \sum fy^2/80 - \bar{y}^2 = 544/80 - 0.25 = 6.8 - 0.25 = 6.55$ M1

$x = 5y + 35$, so $\sigma_x = 5\sigma_y = 5\sqrt{6.55} = 12.79\ldots$ M1

$\boxed{\text{standard deviation} = 12.8 \text{ years}}$ A1

(Total for Question 2 is 13 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

3.

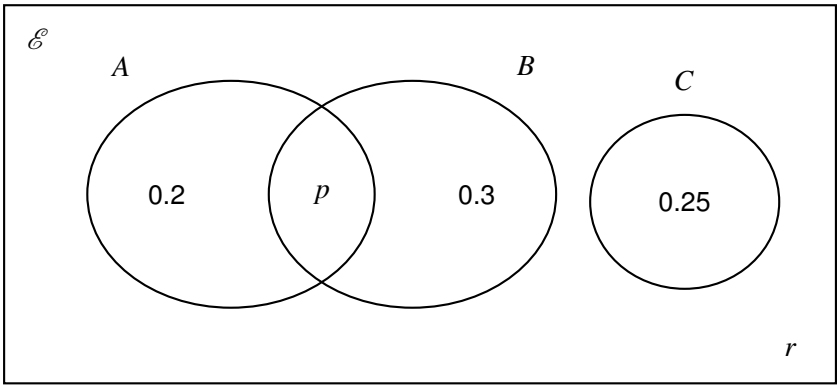


Figure 1

The Venn diagram in Figure 1 shows the probabilities associated with three events A, B and C, where p and r are probabilities.

The events A and B are independent.

(a) Find the value of p, explaining why any other possible value must be rejected.

(5)

(b) Find the value of r

(1)

(c) Find $\mathrm{P}\left(A \mid A \cup B\right)$

(2)

(d) Find $\mathrm{P}\left(B' \mid A'\right)$

(2)

The experiment is carried out twice, independently.

(e) Find the probability that the event $A \cup B$ occurs exactly once.

(2)

(a) $P(A) = 0.2 + p$, $P(B) = 0.3 + p$ M1

Independent: $P(A \cap B) = P(A)P(B) \Rightarrow p = (0.2 + p)(0.3 + p)$ M1

$p = 0.06 + 0.5p + p^2 \Rightarrow p^2 - 0.5p + 0.06 = 0$ A1

$(p - 0.2)(p - 0.3) = 0 \Rightarrow p = 0.2$ or $p = 0.3$ A1

$r = 1 - (0.2 + p + 0.3 + 0.25) = 0.25 - p$: $p = 0.3$ gives $r = -0.05 < 0$ $\boxed{p = 0.2}$ B1

(b) $r = 0.25 - 0.2$ $\boxed{r = 0.05}$ B1

(c) $P(A) = 0.4$, $P(A \cup B) = 0.2 + 0.2 + 0.3 = 0.7$; $P(A \mid A \cup B) = 0.4/0.7$ M1

(A is inside $A \cup B$, so $P(A \cap (A \cup B)) = P(A)$) $\boxed{4/7}$ A1

(d) Neither A nor B: $0.25 + 0.05 = 0.3$; $P(A') = 0.6$; $P(B' \mid A') = 0.3/0.6$ M1

$\boxed{0.5}$ (= $P(B')$, as A and B are independent ✓) A1

DO NOT WRITE IN THIS AREA

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

Question 3 continued

(e) Occurs (0.7) then not (0.3), or not then occurs: $2 \times 0.7 \times 0.3$ 0.42 M1 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

(Total for Question 3 is 12 marks)

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

4.

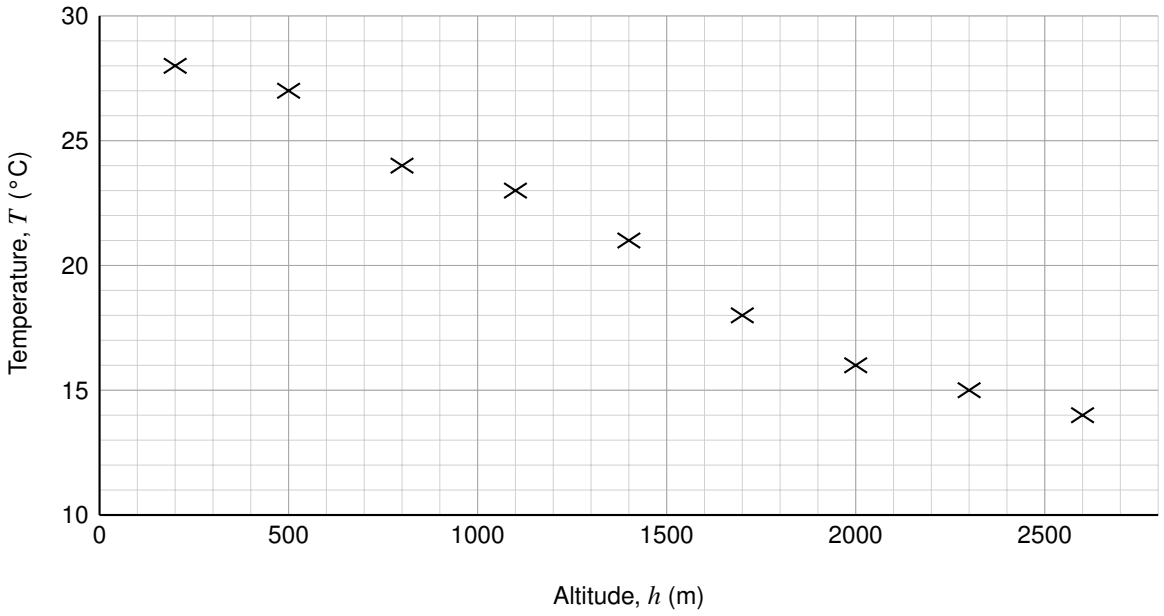


Figure 2

Figure 2 shows the altitude, h metres, and the mean daytime temperature, T °C, at 9 weather stations on a mountain.

The data are coded using

$$p = \frac{h - 1000}{100} \quad \text{and} \quad q = \frac{30 - T}{2}$$

and the coded results are summarised below.

$$\sum p = 36 \quad \sum q = 42 \quad \sum p^2 = 684 \quad \sum q^2 = 250 \quad \sum pq = 337.5$$

(a) Find S_{pp}, S_{qq} and S_{pq}

(3)

(b) Find the product moment correlation coefficient between h and T

(3)

(c) Find the equation of the regression line of T on h, giving your answer in the form $T = a + bh$

(5)

(d) Interpret the value of b in the context of the question.

(1)

(e) Explain why this regression line should not be used to estimate the altitude of a weather station where $T = 20$

(1)

(a) $S_{pp} = 684 - 36^2/9 = 684 - 144 = 540$; $S_{qq} = 250 - 42^2/9 = 250 - 196 = 54$ M1

$S_{pq} = 337.5 - (36 \times 42)/9 = 337.5 - 168$ $S_{pp} = 540$ $S_{qq} = 54$ A1

DO NOT WRITE IN THIS AREA

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

Question 4 continued

$S_{pq} = 169.5$ A1

(b) $r_{pq} = 169.5 / \sqrt{540 \times 54} = 169.5 / 170.76... = 0.9926...$ M1 A1

$q = (30 - T)/2$ falls as T rises (negative coding), so the sign flips $r = -0.993$ B1

(c) $b_q = 169.5/540 = 0.31389...$; $\bar{p} = 36/9 = 4$, $\bar{q} = 42/9 = 4.6667$ M1

$q = 4.6667 - 0.31389 \times 4 + 0.31389p = 3.4111 + 0.31389p$ M1

$(30 - T)/2 = 3.4111 + 0.31389(h - 1000)/100$ M1

$30 - T = 6.8222 + 0.0062778h - 6.2778 = 0.5444 + 0.0062778h$ M1

$T = 29.5 - 0.00628h$ A1

(d) The temperature falls by about 0.628 °C for every 100 m higher up. B1

(e) h is the explanatory variable; estimating h from T needs the line of h on T. B1

(Total for Question 4 is 13 marks)

DO NOT WRITE IN THIS AREA

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

5. The masses, X grams, of eggs from a farm are modelled by a normal distribution with mean μ and standard deviation σ

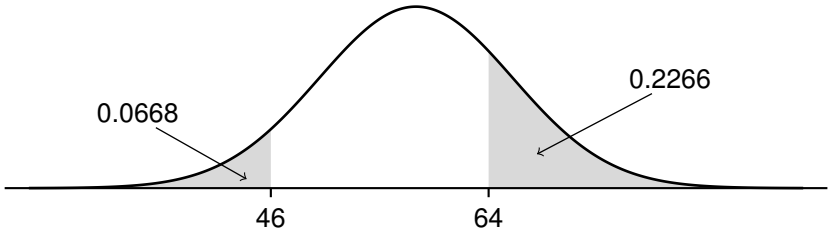


Figure 3

Figure 3 shows that P($X < 46$) = 0.0668 and P($X > 64$) = 0.2266

(a) Find the value of μ and the value of σ

(5)

(b) Find the value of k such that P $\left(|X - \mu| < k\right)$ = 0.9

(3)

The farm records the mass of each egg to the nearest gram.

(c) Find the probability that the recorded mass of a randomly chosen egg is 60 grams.

(2)

(d) Given that the recorded mass of an egg is at least 60 grams, find the probability that its true mass is more than 70 grams.

(2)

(a) 0.0668 is not a percentage point: $1 - 0.0668 = 0.9332 = P(Z < 1.5)$, so $z = -1.5$ B1

$1 - 0.2266 = 0.7734 = P(Z < 0.75)$, so $z = 0.75$ B1

$(46 - \mu)/\sigma = -1.5$ and $(64 - \mu)/\sigma = 0.75 \Rightarrow \mu - 1.5\sigma = 46$, $\mu + 0.75\sigma = 64$ M1

Subtract: $2.25\sigma = 18$ M1

$\sigma = 8$, $\mu = 46 + 12$ $\boxed{\mu = 58}$ $\boxed{\sigma = 8}$ A1

(b) $|X - 58| < k$ is symmetrical: 0.05 in each tail, so $P(X < 58 + k) = 0.95$ M1

$k/8 = 1.6449 \Rightarrow k = 13.159...$ M1

$\boxed{k = 13.2}$ A1

(c) Recorded 60 means $59.5 \leqslant X < 60.5$: z from 0.1875 to 0.3125, so $P(0.19 < Z < 0.31)$ M1

$= 0.6217 - 0.5753$ $\boxed{0.0464}$ A1

(d) Recorded at least 60 means $X \geqslant 59.5$. $P(X > 70) = P(Z > 1.5) = 0.0668$ M1

$P(X > 70 \mid X \geqslant 59.5) = 0.0668 / (1 - 0.5753) = 0.0668/0.4247$ $\boxed{0.157}$ A1

(Total for Question 5 is 12 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

6.

Key: 8 | 2 | 6 means 28 seconds for group A and 26 seconds for group B

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

Figure 4

The back-to-back stem and leaf diagram in Figure 4 shows the times, in seconds, taken by two groups of 15 students to solve a puzzle.

(a) Find the median and the quartiles of the times for each group.

(3)

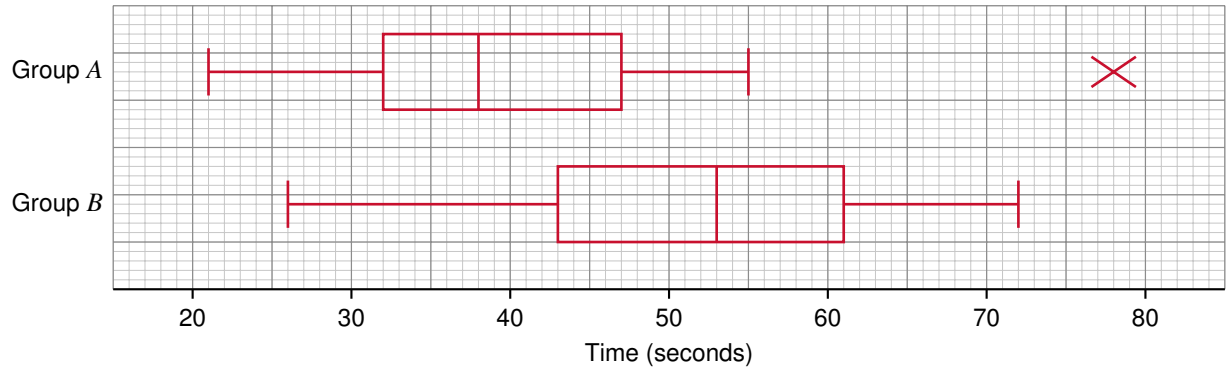
An outlier is a value greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$

(b) Show that group A has exactly one outlier and that group B has none.

(2)

(c) On the grid below, draw box plots for the two groups.

(3)



(d) Compare the times of the two groups.

(2)

(e) For group A, $\sum x = 607$ and $\sum x^2 = 27\,331$

Evaluate $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$ for group A, and use it to comment on whether a normal distribution would be a suitable model for these times.

(3)

(a) $n = 15$: Q_1, Q_2, Q_3 are the 4th, 8th, 12th values. Left leaves: 8 | 2 is 28. M1

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 6 continued

Group A: 21, 25, 28, 32, 33, 35, 36, 38, ... A: $Q_1 = 32$, $Q_2 = 38$, $Q_3 = 47$ A1

B: $Q_1 = 43$, $Q_2 = 53$, $Q_3 = 61$ A1

(b) A: IQR 15, limits $32 - 22.5 = 9.5$ and $47 + 22.5 = 69.5$. B: IQR 18, limits 16 and 88 M1

A: only $78 > 69.5$ is an outlier. B: all values from 26 to 72 lie inside 16 to 88 ✓ A1

(c) Drawn on the grid above. A: box 32, 38, 47 B1

B: box 43, 53, 61, whiskers to 26 and 72 B1

A: whiskers to 21 and 55 (largest non-outlier), cross at 78 B1

(d) Median B (53) > median A (38): group B generally took longer. B1

B has the larger IQR (18 against 15), but A is positively skewed with an outlier. B1

(e) mean $= 607/15 = 40.47$; s.d. $= \sqrt{27331/15 - 40.47^2} = \sqrt{184.52} = 13.58$ M1

$3(40.47 - 38)/13.58 = 0.545$ M1

0.545: positive skew, but a normal model is symmetrical, so it is not suitable. A1

(Total for Question 6 is 13 marks)

TOTAL FOR PAPER IS 75 MARKS

END

DO NOT WRITE IN THIS AREA

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