MODEL ANSWERS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International
Advanced Level

Centre Number | Candidate Number

Thursday 17 January 2019

Afternoon (Time: 1 hour 30 minutes) | Paper Reference **WST01/01**

Statistics S1

Advanced/Advanced Subsidiary

You must have:
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B). Coloured pencils and highlighter pens must not be used.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided
  – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- The total mark for this paper is 75.
- The marks for **each** question are shown in brackets
  – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

P54879A

1/1/1/

P 5 4 8 7 9 A 0 1 2 0

Pearson

Leave blank

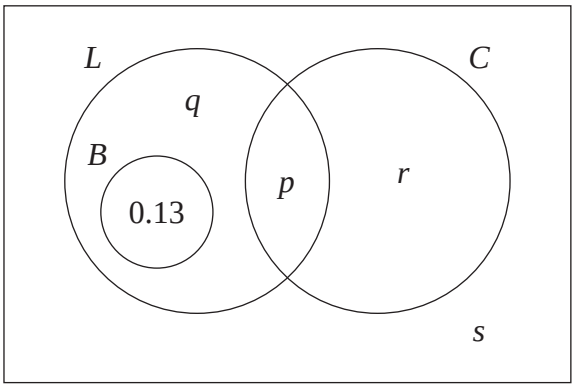
1. The Venn diagram shows the probability of a randomly selected student from a school being in the sets L, B and C, where

L represents the event that the student has instrumental music lessons

B represents the event that the student plays in the school band

C represents the event that the student sings in the school choir

p, q, r and s are probabilities.



(a) Select a pair of mutually exclusive events from L, B and C.

(1)

Given that P(L) = 0.4, P(B) = 0.13, P(C) = 0.3 and the events L and C are independent,

(b) find the value of p,

(2)

(c) find the value of q, the value of r and the value of s.

(3)

A student is selected at random from those who play in the school band or sing in the school choir.

(d) Find the exact probability that this student has instrumental music lessons.

(3)

a) B & C

b) $p = P(L \cap C)$
$= P(L) \times P(C)$
$= 0.4 \times 0.3$
$= 0.12$

c) $p + q + 0.13 = 0.4$
$q = 0.15$

Leave blank

Question 1 continued

$p + r = 0.3$

$r = 0.18$

$s = 1 - (0.4 + 0.3 - p)$

$= 0.42$

d) $P(L \mid B \cup C) = \dfrac{P(L \cap (B \cup C))}{P(B \cup C)}$

$= \dfrac{0.13 + 0.12}{0.13 + 0.3}$

$= \dfrac{0.25}{0.43}$

$= \dfrac{25}{43}$

Q1

(Total 9 marks)

Leave blank

2. The discrete random variable X has the following probability distribution.

x	−2	−1	0	1	3
$P(X = x)$	0.15	a	b	a	0.4

(a) Find $E(X)$.

(2)

Given that $E(X^2) = 4.54$

(b) find the value of a and the value of b.

(5)

The random variable $Y = 3 - 2X$

(c) Find $\text{Var}(Y)$.

(3)

a) $E(X) = -2(0.15) + (-1)a + 0b + (1)a + 3(0.4)$

$= -0.3 + 1.2$

$= 0.9$

b) $E(X^2) = (-2)^2(0.15) + (-1)^2a + (1)^2a + 3^2(0.4)$

$= 4.2 + 2a$

$4.2 + 2a = 4.54$

$a = 0.17$

$0.15 + 2a + b + 0.4 = 1$

$b = 0.11$

c) $\text{Var}(X) = E(X^2) - [E(X)]^2$

$= 4.54 - 0.9^2$

$= 3.73$

$\text{Var}(Y) = (-2)^2(3.73)$

$= 14.92$

Leave blank

Question 2 continued

(Total 10 marks)

Q2

Leave blank

3. The weights of women boxers in a tournament are normally distributed with mean 64 kg and standard deviation 8 kg.

(a) Find the probability that a randomly chosen woman boxer in the tournament weighs less than 51 kg.

(3)

In the tournament, women boxers who weigh less than 51 kg are classified as lightweight. Ren weighs 49 kg and she has a match against another randomly selected, lightweight woman boxer.

(b) Find the probability that Ren weighs less than the other boxer.

(4)

In the tournament, women boxers who weigh more than H kg are classified as heavyweight. Given that 10% of the women boxers in the tournament are classified as heavyweight,

(c) find the value of H.

(3)

a) $W \sim N(64, 8^2)$

$P(W < 51) = P\left(Z < \frac{51-64}{8}\right)$

$= P(Z < -1.625)$

$= 1 - 0.9484$

$= 0.0516$

b) $P(W > 49 \mid W < 51) = \frac{P(49 < W < 51)}{P(W < 51)}$

$P(W < 49) = P\left(Z < \frac{49-64}{8}\right)$

$= P(Z < -1.875)$

$= 1 - 0.9699$

$= 0.0301$

$P(49 < W < 51) = P(W < 51) - P(W < 49)$

$= 0.0516 - 0.0301$

$= 0.0215$

$P(W > 49 \mid W < 51) = \frac{0.0215}{0.0516}$

$= 0.4166...$

≈ 0.417

Leave blank

Question 3 continued

c) $P(W > H) = 0.10$

$P(z > 1.2816) = 0.10$

$\frac{H - 64}{8} = 1.2816$

$H = 74.2528$

≈ 74.3

Q3

(Total 10 marks)

Leave blank

4. A group of 100 adults recorded the amount of time, t minutes, they spent exercising each day. Their results are summarised in the table below.

Time (t minutes)	Frequency (f)	Time midpoint (x)
$0 \leqslant t < 15$	25	7.5
$15 \leqslant t < 30$	17	22.5
$30 \leqslant t < 60$	28	45
$60 \leqslant t < 120$	24	90
$120 \leqslant t \leqslant 240$	6	180

C.F.
25
42
70

[You may use $\sum fx^2 = 455\,512.5$]

A histogram is drawn to represent these data.

The bar representing the time $0 \leqslant t < 15$ has width 0.5 cm and height 6 cm.

(a) Calculate the width and height of the bar representing a time of $60 \leqslant t < 120$

(3)

(b) Use linear interpolation to estimate the median time spent exercising by these adults each day.

(2)

(c) Find an estimate of the mean time spent exercising by these adults each day.

(2)

(d) Calculate an estimate for the standard deviation of these times.

(2)

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

(1)

Further analysis of the above data revealed that 18 of the 25 adults in the $0 \leqslant t < 15$ group took no exercise each day.

(f) State, giving a reason, what effect, if any, this new information would have on your answers to

(i) the estimate of the median in part (b),

(ii) the estimate of the mean in part (c),

(iii) the estimate of the standard deviation in part (d).

(3)

Leave blank

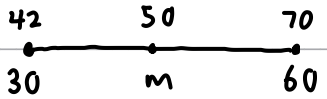
Question 4 continued

a) width : $\frac{60}{15}(0.5) = 2$ cm

area : $\frac{24}{25}(10.5)(6) = \frac{72}{25}$ cm^2

height : $\frac{72}{25} \div 2 = 1.44$ cm

b)



$\frac{M-30}{50-42} = \frac{60-30}{70-42}$

$= \frac{15}{14}(8) + 30$

$= \frac{270}{7}$

≈ 38.6

c) $\Sigma fx = (25)(7.5) + (17)(22.5) + (28)(45) + (24)(90) + (6)(180)$

$= 5070$

$\frac{\Sigma fx}{n} = \frac{5070}{100}$

$= 50.7$

d) $\sigma = \sqrt{\frac{\Sigma fx^2}{n} - \left(\frac{\Sigma fx}{n}\right)^2}$

$= \sqrt{\frac{455\,512.5}{100} - (50.7)^2}$

$= \sqrt{1984.635}$

$= 44.549$

≈ 44.5

e) mean > median

$\therefore$ positive skew

f i) No change because all 18 values are still below median

ii) Decrease. fx will be lower so Σfx decrease

iii) Increase as data becomes more spread out.

Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 13 marks)

Q4

Leave blank

5. Some children are playing a game involving throwing a ball into a bucket. Each child has 3 throws and the number of times the ball lands in the bucket, x, is recorded. Their results are given in the table below.

x	0	1	2	3
Frequency	16	36	24	4

(a) Find $\bar{x}$

(1)

Sandra decides to model the game by assuming that on each throw, the probability of the ball landing in the bucket is 0.4 for every child on every throw and that the throws are all independent. The random variable S represents the number of times the ball lands in the bucket for a randomly selected child.

(b) Find P(S = 2)

(2)

(c) Complete the table below to show the probability distribution for S.

s	0	1	2	3
P($S = s$)	0.216	0.432	0.288	0.064

(1)

Ting believes that the probability of the ball landing in the bucket is not the same for each throw. He suggests that the probability will increase with each throw and uses the model

$$p_i = 0.15i + 0.10$$

where i = 1, 2, 3 and p_i is the probability that the ith throw of the ball, by any particular child, will land in the bucket.
The random variable T represents the number of times the ball lands in the bucket for a randomly selected child using Ting's model.

(d) Show that

(i) P(T = 3) = 0.055

(ii) P(T = 1) = 0.45

(5)

(e) Complete the table below to show the probability distribution for T, stating the exact probabilities in each case.

t	0	1	2	3
P($T = t$)	0.2025	0.45	0.2925	0.055

(3)

(f) State, giving your reasons, whether Sandra's model or Ting's model is the more appropriate for modelling this game.

(3)

Leave blank

Question 5 continued

a) $\bar{n} = \frac{16(0) + 36(1) + 24(2) + 3(4)}{16 + 36 + 24 + 4}$

$= \frac{96}{80}$

$= 1.2$

b) $P(S = 2) = 3 \times 0.6 \times 0.4 \times 0.4$

$= 0.288$

c) $(0.6)^3 = 0.216$

$P(S = 0) = 0.216$

di) $p_1 = 0.15 + 0.10$

$= 0.25$

$p_2 = 0.15(2) + 0.10$

$= 0.4$

$p_3 = 0.15(3) + 0.10$

$= 0.55$

$P(T = 3) = 0.25 \times 0.4 \times 0.55$

$= 0.055$

ii) $P(T = 1) = 0.25(0.6)(0.45) + 0.75(0.4)(0.45) + 0.75(0.6)(0.55)$

$= 0.0675 + 0.135 + 0.2475$

$= 0.45$

e) $P(T = 0) = (0.75)(0.6)(0.45)$

$= 0.2025$

$P(T = 2) = 1 - 0.2025 - 0.45 - 0.055$

$= 0.2925$

f) $n = 0 \quad \frac{16}{80} = 0.2$

$n = 1 \quad \frac{36}{80} = 0.45$

$n = 2 \quad \frac{24}{80} = 0.3$

$n = 3 \quad \frac{4}{80} = 0.05$

Ting's model is always closer

$\therefore$ choose Ting's model

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 15 marks)

Q5

Leave blank

6. Following some school examinations, Chetna is studying the results of the 16 students in her class. The mark for paper 1, x, and the mark for paper 2, y, for each student are summarised in the following statistics.

$$\bar{x} = 35.75 \qquad \bar{y} = 25.75 \qquad \sigma_x = 7.79 \qquad \sigma_y = 11.91 \qquad \sum xy = 15\,837$$

(a) Comment on the differences between the marks of the students on paper 1 and paper 2

(2)

Chetna decides to examine these data in more detail and plots the marks for each of the 16 students on the scatter diagram opposite.

(b) (i) Explain why the circled point (38, 0) is possibly an outlier.

(ii) Suggest a possible reason for this result.

(2)

Chetna decides to omit the data point (38, 0) and examine the other 15 students' marks.

(c) Find the value of $\bar{x}$ and the value of $\bar{y}$ for these 15 students.

(3)

For these 15 students

(d) (i) explain why $\sum xy$ is still 15 837

(ii) show that $S_{xy} = 1169.8$

(3)

For these 15 students, Chetna calculates $S_{xx} = 965.6$ and $S_{yy} = 1561.7$ correct to 1 decimal place.

(e) Calculate the product moment correlation coefficient for these 15 students.

(2)

(f) Calculate the equation of the line of regression of y on x for these 15 students, giving your answer in the form $y = a + bx$

(4)

The product moment correlation coefficient between x and y for all 16 students is 0.746

(g) Explain how your calculation in part (e) supports Chetna's decision to omit the point (38, 0) before calculating the equation of the linear regression line.

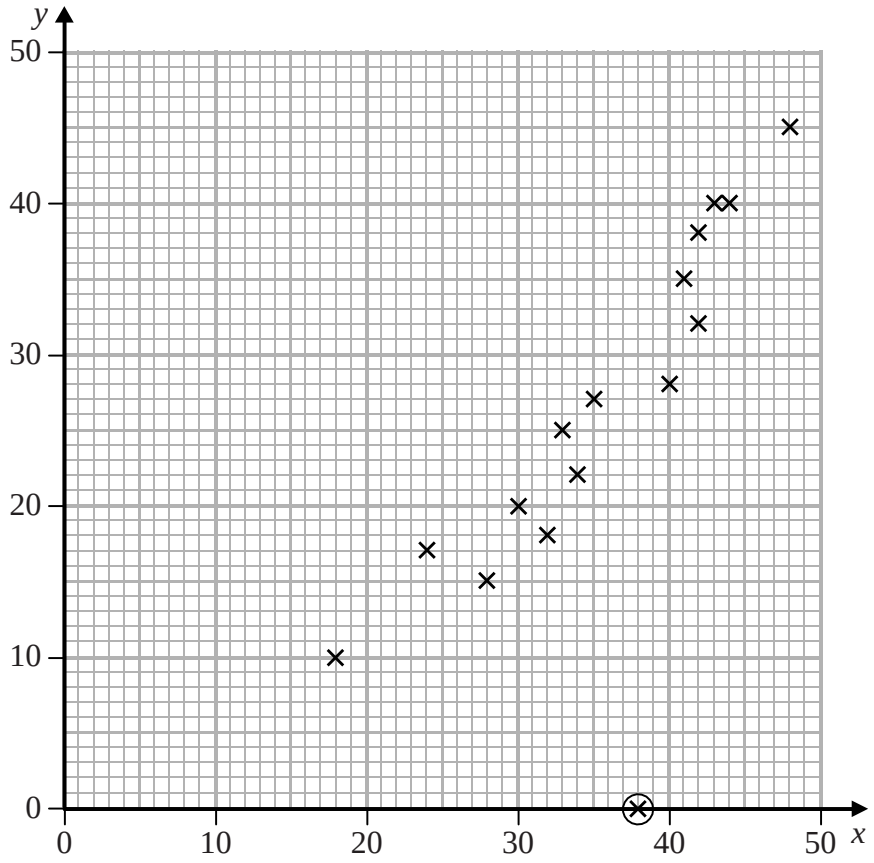
(1)

(h) Estimate the mark in the second paper for a student who scored 38 marks in the first paper.

(1)

Leave blank

Question 6 continued



a) The mean score for paper 1 is higher. The spread of scores for paper 2 is larger.

bi) The point lies far away from the line of best fit.

ii) The student did not sit for paper 2.

c) $\bar{x} = \frac{35.75 \times 16 - 38}{15}$

$= 35.6$

$\bar{y} = 25.75 \times \frac{16}{15}$

$= 27.466...$

$\simeq 27.5$

di) new $\Sigma xy = 15837 - 38(0)$

$= 15837 - 0$

$\therefore$ no change

ii) $\Sigma x = (35.75)(16) - 38$

$= 534$

Leave blank

Question 6 continued

$\Sigma y = 25.75 \times 16$

$= 412$

$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$

$= 15837 - \frac{(534)(412)}{15}$

$= 1169.8$

e) $r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$

$= \frac{1169.8}{\sqrt{965.6 \times 1561.7}}$

$= 0.952607...$

≈ 0.953

f) $b = \frac{S_{xy}}{S_{xx}}$

$= \frac{1169.8}{965.6}$

$= 1.21147$

≈ 1.2

$27.5 = a + 1.2\ (35.6)$

$a = -15.576$

≈ -15.6

$y = -15.6 + 1.2x$

g) $0.953 > 0.746$, there is a stronger linear correlation

h) $y = -15.6 + 1.2\ (38)$

$= 30$

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 18 marks)

TOTAL FOR PAPER: 75 MARKS

END