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 | | | | |

Wednesday 17 October 2018

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

Statistics S1

Advanced/Advanced Subsidiary

You must have: Scientific calculator
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 ▶

P54942A

1/1/1/1/

P 5 4 9 4 2 A 0 1 2 4

Pearson

Leave blank

1. The heights above sea level (h hundred metres) and the temperatures (t °C) at 12 randomly selected places in France, at 7 am on July 31st, were recorded.
The data are summarised as follows

$$\sum h = 112 \quad \sum t = 136 \quad \sum t^2 = 1828 \quad S_{ht} = -236 \quad S_{hh} = 297$$

(a) Find the value of S_{tt}

(2)

(b) Calculate the product moment correlation coefficient for these data.

(2)

(c) Interpret the relationship between t and h.

(1)

(d) Find an equation of the regression line of t on h.

(3)

At 7 am on July 31st Yinka is on holiday in South Africa. He uses the regression equation to estimate the temperature when the height above sea level is 500 m.

(e) Find the estimated temperature Yinka calculates.

(2)

(f) Comment on the validity of your answer in part (e).

(1)

a) $S_{tt} = \Sigma t^2 - \frac{(\Sigma t)^2}{n}$

$= 1828 - \frac{136^2}{12}$

$= \frac{860}{3}$

$\simeq 287$

b) $r = \frac{S_{ht}}{\sqrt{S_{hh} \times S_{tt}}}$

$= \frac{-236}{\sqrt{297 \times 287}}$

$= \frac{-236}{\sqrt{85140}}$

$= -0.808807...$

$\simeq -0.809$

c) There is a negative relationship between height and temperature. When height increases, temperature decreases.

Leave blank

Question 1 continued

d) $b = \frac{S_{ht}}{S_{hh}}$

$= \frac{-236}{297}$

$= -0.79461$

$\simeq -0.795$

$\bar{t} = a + b\bar{h}$

$\frac{136}{12} = a + (-0.7946)\frac{112}{12}$

$a = 18.7496$

$\simeq 18.7$

$t = 18.7 - 0.795h$

e) $t = 18.7 - 0.795(\frac{500}{100})$

$= 14.776...$

$\simeq 14.8$

f) Unreliable. The data is from France, which is different from South Africa

Leave blank

Question 1 continued

Leave blank

Question 1 continued

(Total 11 marks)

Q1

Leave blank

2. The weights, to the nearest kilogram, of a sample of 33 female spotted hyenas living in the Serengeti are summarised in the stem and leaf diagram below.

Weight (kg)		Totals
3	2 3 7	(3)
4	1 3 3 4 5 5 6 9	(8)
5	1 2 2 3 4 4 5 5 5 7 8 8 9 9 9	(15)
6	2 3 3	(3)
7	1 4 7	(3)
8	4	(1)

Key: 3|2 means 32

(a) Find the median and quartiles for the weights of the female spotted hyenas.

(3)

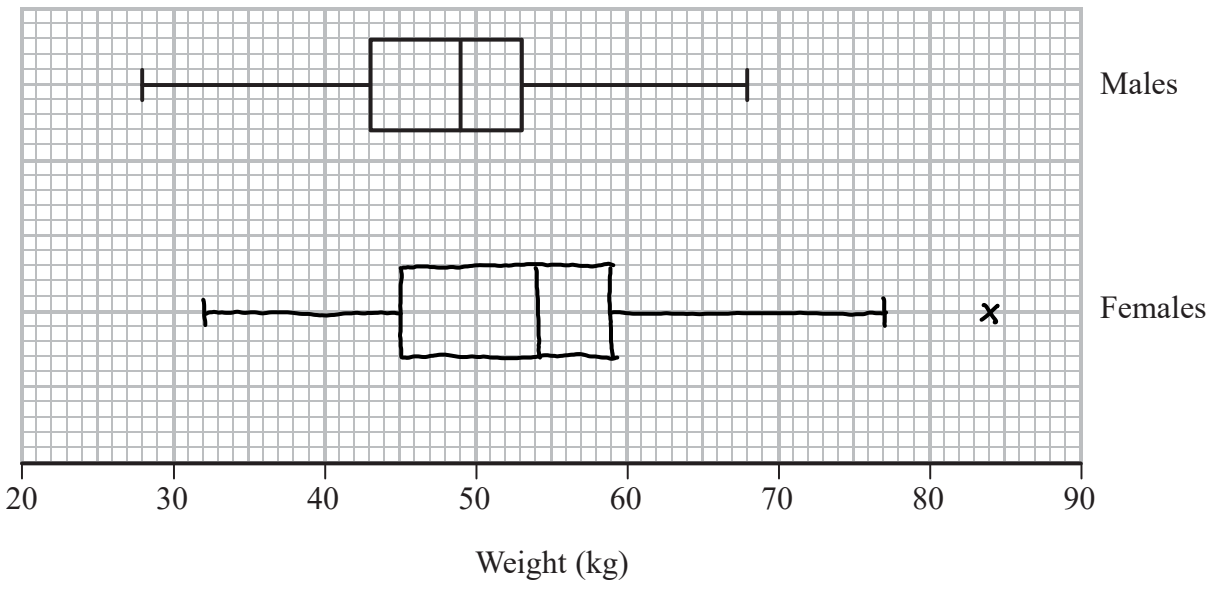
An outlier is defined as any value greater than c or any value less than d where

$$c = Q_3 + 1.5(Q_3 - Q_1)$$
$$d = Q_1 - 1.5(Q_3 - Q_1)$$

(b) Showing your working clearly, identify any outliers for these data.

(3)

The weights, to the nearest kilogram, of a sample of male spotted hyenas living in the Serengeti are summarised below.



(c) In the space provided in the grid above, draw a box and whisker plot to represent the weights of female spotted hyenas living in the Serengeti. Indicate clearly any outliers.
(A copy of this grid is on page 9 if you need to redraw your box and whisker plot.)

(3)

(d) Compare the weights of male and female spotted hyenas living in the Serengeti.

(2)

Leave blank

Question 2 continued

a) $\frac{33}{2} = 16.5$

≈ 17

M = 17th value

= 54

$\frac{33}{4} = 8.25$

Q_1 = 9th value

= 45

$\frac{3(33)}{4} = 24.75$

Q_3 = 25th value

= 59

b) $c = 59 + 1.5(59 - 45)$

= 80

$d = 45 - 1.5(59 - 45)$

= 24

outlier : 84

d) The males have lower median and smaller interquartile range compared to females.

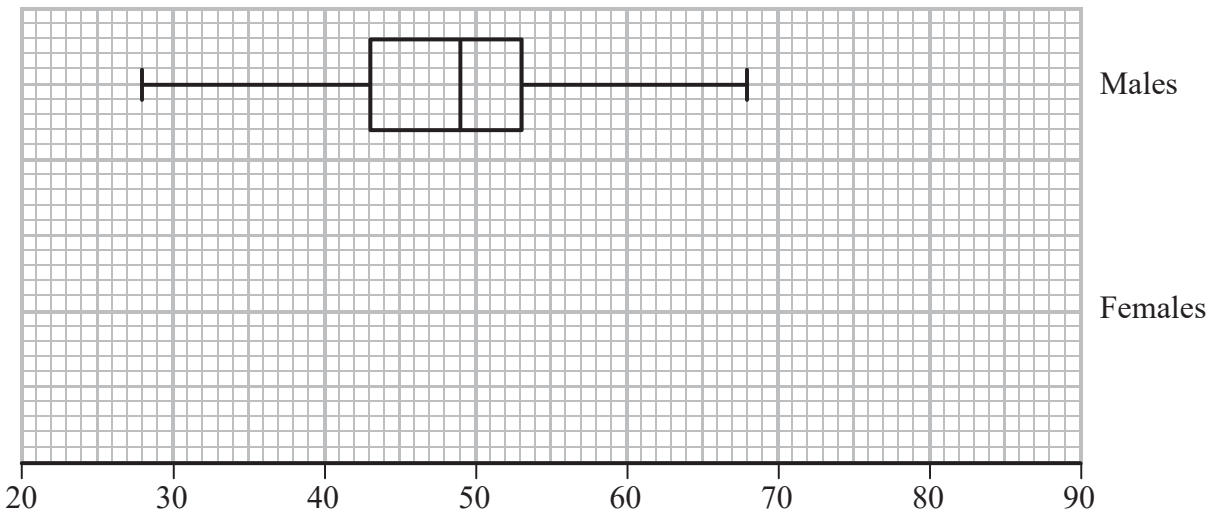
Leave blank

Question 2 continued

Leave blank

Question 2 continued

Only use this grid if you need to redraw your box and whisker plot.



Q2

(Total 11 marks)

Leave blank

3. The parking times, t hours, for cars in a car park are summarised below.

Time (t hours)	Frequency (f)	Time midpoint (m)	c.f.
$0 \leqslant t < 1$	10	0.5	10
$1 \leqslant t < 2$	18	1.5	28
$2 \leqslant t < 4$	15	3	43
$4 \leqslant t < 6$	12	5	55
$6 \leqslant t < 12$	5	9	60

(You may use $\sum fm = 182$ and $\sum fm^2 = 883$)

A histogram is drawn to represent these data.

The bar representing the time $1 \leqslant t < 2$ has a width of 1.5 cm and a height of 6 cm.

(a) Calculate the width and the height of the bar representing the time $4 \leqslant t < 6$

(3)

(b) Use linear interpolation to estimate the median parking time for the cars in the car park.

(2)

(c) Estimate the mean and the standard deviation of the parking time for the cars in the car park.

(3)

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

(2)

One of these cars is selected at random.

(e) Estimate the probability that this car is parked for more than 75 minutes.

(3)

a) width: $\frac{6-4}{2-1} \times 1.5 = 3$ cm

area: $\frac{12}{18} \times (1.5 \times 6) = 6\ cm^2$

height: $\frac{6}{3} = 2$ cm

b) $10 + 18 + 15 + 12 + 5 = 60$

$\frac{60}{2} = 30$

28 — 30 — 43

2 — M — 4

$\frac{30-28}{M-2} = \frac{43-28}{4-2}$

$M - 2 = \frac{2}{7.5}$ $\quad M = \frac{34}{15}$

≈ 2.27

Leave blank

Question 3 continued

c) $\bar{t} = \frac{\Sigma fm}{n}$

$= \frac{182}{60}$

$= 3.0333$

≈ 3.03

$\sigma = \sqrt{\frac{\Sigma fm^2}{n} - \bar{t}^2}$

$= \sqrt{\frac{883}{60} - 3.033^2}$

$= 2.34852$

≈ 2.35

d) Positive skew as mean > median

e) $\frac{75}{60} = 1.25$

$P(T < 1.25) = (10 + \frac{1}{4} \times 18) \div 60$

$= \frac{29}{120}$

$P(T > 1.25) = 1 - \frac{29}{120}$

$= \frac{91}{120}$

≈ 0.758

Leave blank

Question 3 continued

Leave blank

Question 3 continued

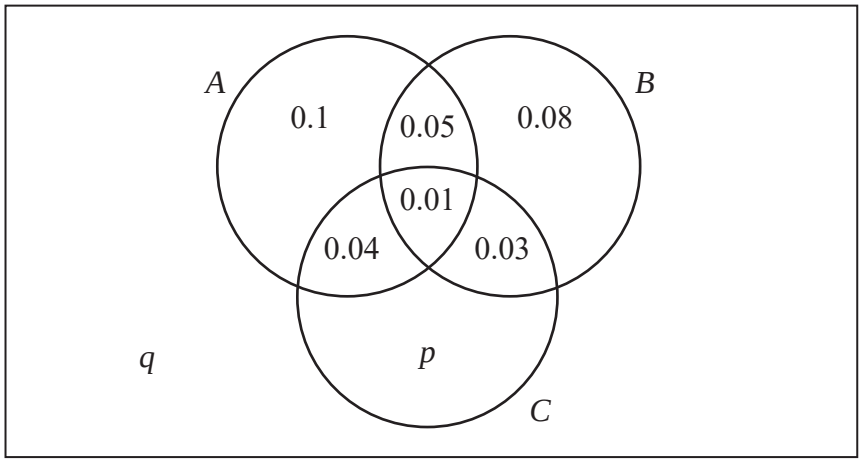
(Total 13 marks)

Q3

Leave blank

4. Pieces of wood cladding are produced by a timber merchant. There are three types of fault, *A*, *B* and *C*, that can appear in each piece of wood cladding.

The Venn diagram shows the probabilities of a piece of wood cladding having the various types of fault.



A piece of wood cladding is chosen at random.

(a) Find the probability that the piece of wood cladding has more than one type of fault.

(1)

Fault types *A* and *C* occur independently.

(b) Find the probability that the piece of wood cladding has no faults.

(4)

Given that the piece of wood cladding has fault *A*,

(c) find the probability that it also has fault *B* but not fault *C*.

(2)

Two pieces of the wood cladding are selected at random.

(d) Find the probability that both have exactly 2 types of fault.

(3)

a) $0.01 + 0.03 + 0.04 + 0.05 = 0.13$

b) $P(A) \times P(C) = P(A \cap C)$

$(0.1 + 0.01 + 0.04 + 0.05) \times (0.01 + 0.03 + 0.04 + p) = (0.04 + 0.01)$

$0.2 \times (0.08 + p) = 0.05$

$p = 0.17$

$q = 1 - 0.2 - 0.17 - 0.03 - 0.08 = 0.52$

Leave blank

Question 4 continued

c) $P(B \cap C' | A) = \frac{0.05}{0.2}$

$= 0.25$

d) $P(\text{exactly 2 defects}) = 0.03 + 0.04 + 0.05$

$= 0.12$

$0.12 \times 0.12 = 0.0144$

Q4

(Total 10 marks)

Leave blank

5. The discrete random variable X is defined by the cumulative distribution function

x	1	2	3	4	5
F(x)	$\frac{3k}{2}$	$4k$	$\frac{15k}{2}$	$12k$	$\frac{35k}{2}$

where k is a constant.

(a) Find the probability distribution of X.

(3)

(b) Find $P(1.5 < X \leqslant 3.5)$

(2)

The random variable $Y = 12 - 7X$

(c) Calculate Var(Y)

(6)

(d) Calculate $P(4X \leqslant |Y|)$

(3)

a) $\frac{35}{2}k = 1$

$k = \frac{2}{35}$

n	1	2	3	4	5
$P(X=n)$	$\frac{3}{35}$	$\frac{1}{7}$	$\frac{1}{5}$	$\frac{9}{35}$	$\frac{11}{35}$

b) $\frac{1}{7} + \frac{1}{5} = \frac{12}{35}$

c) $E(X) = \frac{3}{35} + \frac{2}{7} + \frac{3}{5} + \frac{36}{35} + \frac{55}{35}$

$= \frac{25}{7}$

$E(X^2) = \frac{3}{35} + (2)^2(\frac{1}{7}) + (3)^2(\frac{1}{5}) + (4)^2(\frac{9}{35}) + (5)^2(\frac{11}{35})$

$= \frac{101}{7}$

$Var(X) = \frac{101}{7} - (\frac{25}{7})^2$

$= \frac{82}{49}$

$Var(Y) = 7^2 \times Var(X)$

$= 49 \times \frac{82}{49}$

$= 82$

Leave blank

Question 5 continued

d) $P(4X \leq |Y|) = P(4X \leq |12-7X|)$

$= P(4X \leq 12-7X) + P(4X \leq -12+7X)$

$= P(X \leq \frac{12}{11}) + P(X \geqslant 4)$

$= \frac{3+9+11}{35}$

$= \frac{23}{35}$

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 14 marks)

Q5

Leave blank

6. A machine makes bolts such that the length, L cm, of a bolt has distribution $L \sim N(4.1, 0.125^2)$

A bolt is selected at random.

(a) Find the probability that the length of this bolt is more than 4.3 cm.

(3)

(b) Show that $P(3.9 < L < 4.3)$ is 0.890 correct to 3 decimal places.

(1)

The machine makes 500 bolts.
The cost to make each bolt is 5 pence.
Only bolts with length between 3.9 cm and 4.3 cm can be used.
These are sold for 9 pence each.
All the bolts that cannot be used are recycled with a scrap value of 1 pence each.

(c) Calculate an estimate for the profit made on these 500 bolts.

(4)

Following adjustments to the machine, the length of a bolt, B cm, made by the machine is such that $B \sim N(\mu, \sigma^2)$

Given that $P(B > 4.198) = 0.025$ and $P(B < 4.065) = 0.242$

(d) find the value of μ and the value of σ

(6)

(e) State, giving a reason, whether the adjustments to the machine will result in a decrease or an increase in the profit made on 500 bolts.

(2)

a) $P(L > 4.3) = P\left(Z > \frac{4.3-4.1}{0.125}\right)$

$= P(Z > 1.6)$
$= 1 - 0.9452$
$= 0.0548$

b) $P(3.9 < L < 4.3) = P(L < 4.3) - P(L < 3.9)$
$= 0.9452 - P\left(Z < \frac{3.9-4.1}{0.125}\right)$

$= 0.9452 - P(Z < -1.6)$
$= 0.9452 - (1 - 0.9452)$
$= 0.8904$

Leave blank

Question 6 continued

c) Total costs : $500 \times 5 = 2500$

From selling bolts : $0.8904 \times 500 = 445.2$

≈ 445

$445 \times 9 = 4005$

From recycling : $(500 - 445) \times 1 = 55$

$4005 + 55 - 2500 = 1560$ pence

d) $P(B > 4.198) = P\left(z > \frac{4.198 - \mu}{\sigma}\right)$ $\quad P(z > 1.96) = 0.025$

$\frac{4.198 - \mu}{\sigma} = 1.96$ — ①

$P(B < 4.065) = P\left(z < \frac{4.065 - \mu}{\sigma}\right)$ $\quad P(z < -0.7) = 0.242$

$\frac{4.065 - \mu}{\sigma} = -0.7$ — ②

① - ② $\quad \frac{0.133}{\sigma} = 2.66$

$\sigma = 0.05$

$\frac{4.065 - \mu}{0.05} = -0.7$

$\mu = 4.1$

e) Mean is the same but s.d. is lower so $P(3.9 < L < 4.3)$ increases

∴ profit will increase

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 16 marks)

TOTAL FOR PAPER: 75 MARKS

END