Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

**Pearson Edexcel**
**International Advanced Level**

Centre Number | Candidate Number

# Friday 12 June 2020

Afternoon (Time: 1 hour 30 minutes) | Paper Reference **WST02/01**

# Mathematics

## International Advanced Subsidiary/Advanced Level
## Statistics S2

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

Total Marks

**Candidates may use any calculator permitted by Pearson regulations. 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).
- **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. If a calculator is used instead of the tables, the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

## Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 6 questions in this question paper. 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.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.

*Turn over* ▶

P65762A

©2020 Pearson Education Ltd.

1/1/1/

P 6 5 7 6 2 A 0 1 2 4

Pearson

Leave blank

**1.**

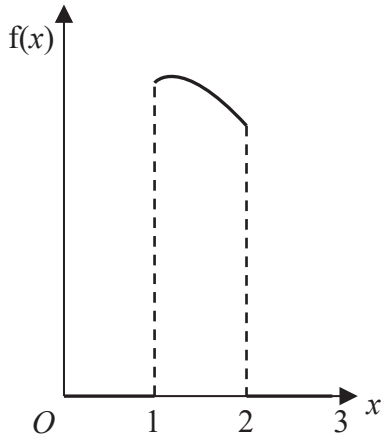



**Figure 1**

Figure 1 shows a sketch of the probability density function f($x$) of the random variable $X$.

For $1 \leqslant x \leqslant 2$, f($x$) is represented by a curve with equation $f(x) = k\left(\frac{1}{2}x^3 - 3x^2 + ax + 1\right)$ where $k$ and $a$ are constants.

For all other values of $x$, f($x$) = 0

(a) Use algebraic integration to show that $k(12a - 33) = 8$

**(4)**

Given that $a = 5$

(b) calculate the mode of $X$.

**(4)**

(a) $\int_1^2 k\left(\frac{1}{2}x^3 - 3x^2 + ax + 1\right) dx = 1$

$k\left[\frac{x^4}{8} - x^3 + \frac{ax^2}{2} + x\right]_1^2 = 1$

$k(4 - 8 + 2a + 2) - k\left(\frac{1}{8} - 1 + \frac{a}{2} + 1\right) = 1$

$k(2a - 4) - k\left(\frac{1}{8} + \frac{a}{2}\right) = 1$

$-\frac{33}{8}k + \frac{3}{2}ka = 1$

$\times 8: \quad -33k + 12ka = 8$

$\therefore \quad k(12a - 33) = 8 \quad$ as required

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

Leave blank

**Question 1 continued**

(b) $a = 5$ : $k(60 - 33) = 8 \Rightarrow k = \frac{8}{27}$

$\frac{df(x)}{dx} = k\left(\frac{3}{2}x^2 - 6x + 5\right)$

At the mode $\frac{df(x)}{dx} = 0$ : $\frac{3}{2}x^2 - 6x + 5 = 0$

$x = \frac{6 \pm \sqrt{(6^2 - 4 \times 1.5 \times 5)}}{3} = \frac{6 \pm \sqrt{6}}{3}$

$x = 2.8164...$ (outside $1 \leq x \leq 2$, so rejected) or $x = 1.1835...$

$\frac{d^2f(x)}{dx^2} = k(3x - 6) < 0$ at $x = 1.18$, so a maximum

Mode $= 2 - \frac{\sqrt{6}}{3} = 1.18$ (3 s.f.)

Leave blank

**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

Leave blank

**Question 1 continued**

**Q1**

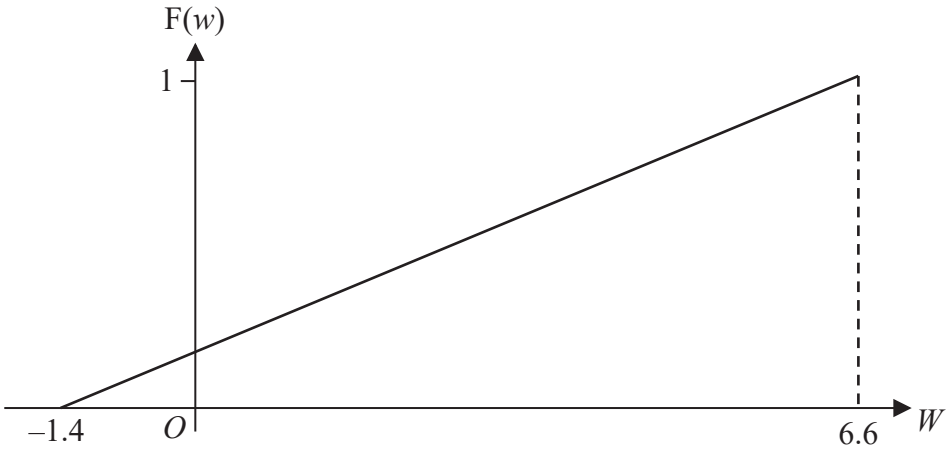
**(Total 8 marks)**

Leave blank

**2.** In the summer Kylie catches a local steam train to work each day. The published arrival time for the train is 10 am.

The random variable $W$ is the train's actual arrival time minus the published arrival time, in minutes. When the value of $W$ is positive, the train is late.

The cumulative distribution function F($w$) is shown in the sketch below.




(a) Specify fully the probability density function f($w$) of $W$.

**(2)**

(b) Write down the value of E($W$)

**(1)**

(c) Calculate $\alpha$ such that $P(\alpha \leqslant W \leqslant 1.6) = 0.35$

**(2)**

A day is selected at random.

(d) Calculate the probability that on this day the train arrives between 1.2 minutes late and 2.4 minutes late.

**(2)**

Given that on this day the train was between 1.2 minutes late and 2.4 minutes late,

(e) calculate the probability that it was more than 2 minutes late.

**(2)**

A random sample of 40 days is taken.

(f) Calculate the probability that for at least 10 of these days the train is between 1.2 minutes late and 2.4 minutes late.

**(3)**

(a) F(w) is linear from −1.4 to 6.6, so W ~ U[−1.4, 6.6]

6.6 − (−1.4) = 8

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

Leave blank

**Question 2 continued**

$$f(w) = \begin{cases} \frac{1}{8} & -1.4 \leq w \leq 6.6 \\ 0 & \text{otherwise} \end{cases}$$

(b) $E(W) = \frac{-1.4 + 6.6}{2} = 2.6$

(c) $P(\alpha \leq W \leq 1.6) = (1.6 - \alpha) \times \frac{1}{8} = 0.35$

$1.6 - \alpha = 2.8$

$\alpha = -1.2$

(d) $P(1.2 < W < 2.4) = (2.4 - 1.2) \times \frac{1}{8}$

$= \frac{3}{20} = 0.15$

(e) $P(W > 2 \mid 1.2 < W < 2.4) = \frac{P(2 < W < 2.4)}{P(1.2 < W < 2.4)}$

$= \frac{0.4 \times 1/8}{0.15} = \frac{0.05}{0.15}$

$= \frac{1}{3}$

(f) $Y$ = number of days between 1.2 and 2.4 minutes late

$Y \sim B(40, 0.15)$

$P(Y \geq 10) = 1 - P(Y \leq 9)$

$= 1 - 0.9328$

$= 0.0672$

Leave blank

**Question 2 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

Leave blank

**Question 2 continued**

**(Total 12 marks)**

**Q2**

Leave blank

**3.** A manufacturer produces plates. The proportion of plates that are flawed is 45%, with flawed plates occurring independently.

A random sample of 10 of these plates is selected.

(a) Find the probability that the sample contains

(i) fewer than 2 flawed plates,

(ii) at least 6 flawed plates.

**(4)**

George believes that the proportion of flawed plates is not 45%. To assess his belief George takes a random sample of 120 plates. The random variable $F$ represents the number of flawed plates found in the sample.

(b) Using a normal approximation, find the maximum number of plates, $c$, and the minimum number of plates, $d$, such that

$$P(F \leqslant c) \leqslant 0.05 \text{ and } P(F \geqslant d) \leqslant 0.05$$

where $F \sim B(120, 0.45)$

**(7)**

The manufacturer claims that, after a change to the production process, the proportion of flawed plates has decreased. A random sample of 30 plates, taken after the change to the production process, contains 8 flawed plates.

(c) Use a suitable hypothesis test, at the 5% level of significance, to assess the manufacturer's claim. State your hypotheses clearly.

**(4)**

(a)(i) $X \sim B(10, 0.45)$

$P(X < 2) = P(X \leq 1) = 0.0233$

(ii) $P(X \geq 6) = 1 - P(X \leq 5)$

$= 1 - 0.7384$

$= 0.2616 \quad = 0.262$

(b) $np = 120 \times 0.45 = 54, \quad np(1-p) = 54 \times 0.55 = 29.7$

$F \approx N(54, 29.7)$

$P(F \leq c) \leq 0.05$ :

$$\frac{c + 0.5 - 54}{\sqrt{29.7}} \leq -1.6449$$

$c \leq 53.5 - 1.6449\sqrt{29.7} = 44.535...$

$c = 44$





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

Leave blank

**Question 3 continued**

$P(F \geqslant d) \leqslant 0.05$ :

$\frac{d - 0.5 - 54}{\sqrt{29.7}} \geqslant 1.6449$

$d \geqslant 54.5 + 1.6449\sqrt{29.7} = 63.464...$

$d = 64$

(c) $H_0 : p = 0.45$ $H_1 : p < 0.45$

$Y$ = number of flawed plates in 30, $Y \sim B(30, 0.45)$ under $H_0$

$P(Y \leqslant 8) = 0.0312$

$0.0312 < 0.05$ so the result is significant: reject $H_0$

(8 lies in the critical region $Y \leqslant 8$)

There is evidence that the proportion of flawed plates has decreased, so the manufacturer's claim is supported.

Leave blank

**Question 3 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

Leave blank

**Question 3 continued**

**(Total 15 marks)**

**Q3**

Leave blank

**4.** In a peat bog, Common Spotted-orchids occur at a mean rate of 4.5 per $m^2$

(a) Give an assumption, not already stated, that is required for the number of Common Spotted-orchids per $m^2$ of the peat bog to follow a Poisson distribution.

**(1)**

Given that the number of Common Spotted-orchids in 1 $m^2$ of the peat bog can be modelled by a Poisson distribution,

(b) find the probability that in a randomly selected 1 $m^2$ of the peat bog

(i) there are exactly 6 Common Spotted-orchids,

(ii) there are fewer than 10 but more than 4 Common Spotted-orchids.

**(4)**

Juan believes that by introducing a new management scheme the number of Common Spotted-orchids in the peat bog will increase. After three years under the new management scheme, a randomly selected 2 $m^2$ of the peat bog contains 11 Common Spotted-orchids.

(c) Using a 5% significance level assess Juan's belief. State your hypotheses clearly.

**(5)**

Assuming that in the peat bog, Common Spotted-orchids still occur at a mean rate of 4.5 per $m^2$

(d) use a normal approximation to find the probability that in a randomly selected 20 $m^2$ of the peat bog there are fewer than 70 Common Spotted-orchids.

**(3)**

Following a period of dry weather, the probability that there are fewer than 70 Common Spotted-orchids in a randomly selected 20 $m^2$ of the peat bog is 0.012

A random sample of 200 non-overlapping 20 $m^2$ areas of the peat bog is taken.

(e) Using a suitable approximation, calculate the probability that at most 1 of these areas contains fewer than 70 Common Spotted-orchids.

**(3)**

(a) The orchids occur randomly (singly and independently of each other) in the peat bog.

(b)(i) $S \sim Po(4.5)$

$P(S = 6) = \frac{e^{-4.5} 4.5^6}{6!} = 0.12812... = 0.128$

(ii) $P(4 < S < 10) = P(S \leq 9) - P(S \leq 4)$

$= 0.9829 - 0.5321$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

$= 0.4508 \quad = 0.451$

(c) $H_0 : \lambda = 9 \qquad H_1 : \lambda > 9$

Under $H_0$, $M \sim Po(9)$ (2 $m^2$ at 4.5 per $m^2$)

$P(M \geq 11) = 1 - P(M \leq 10) = 1 - 0.7060$

$= 0.2940$

$0.294 > 0.05$ so not significant: do not reject $H_0$

(11 is not in the critical region $M \geq 15$)

There is insufficient evidence to support Juan's belief that the number of Common Spotted-orchids has increased.

(d) $T \sim Po(90) \approx N(90, 90)$

$P(T < 70) \approx P\left(Z < \frac{69.5 - 90}{\sqrt{90}}\right)$

$= P(Z < -2.16) \quad = 1 - 0.9846$

$= 0.0154$

(e) $V \sim B(200, 0.012)$, $n$ large and $p$ small

$V \approx Po(200 \times 0.012) = Po(2.4)$

$P(V \leq 1) = P(V = 0) + P(V = 1) = e^{-2.4}(1 + 2.4)$

$= 0.30844... \quad = 0.308$

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

Leave blank

**Question 4 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

Leave blank

**Question 4 continued**

**(Total 16 marks)**

**Q4**

Leave blank

**5.** The waiting time, $T$ minutes, of a customer to be served in a local post office has probability density function

$$f(t) = \begin{cases} \frac{1}{50}(18 - 2t) & 0 \leqslant t \leqslant 3 \\ \frac{1}{20} & 3 < t \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

Given that the mean number of minutes a customer waits to be served is 1.66

(a) use algebraic integration to find Var($T$), giving your answer to 3 significant figures.

**(5)**

(b) Find the cumulative distribution function F($t$) for all values of $t$.

**(4)**

(c) Calculate the probability that a randomly chosen customer's waiting time will be more than 2 minutes.

**(2)**

(d) Calculate P([E($T$) – 2] < $T$ < [E($T$) + 2])

**(2)**

(a) $E(T^2) = \int_0^3 \frac{1}{50}(18t^2 - 2t^3)\,dt + \int_3^5 \frac{1}{20} t^2\,dt$

$= \left[\frac{1}{50}\left(6t^3 - \frac{t^4}{2}\right)\right]_0^3 + \left[\frac{t^3}{60}\right]_3^5$

$= \frac{1}{50}\left(6 \times 3^3 - \frac{3^4}{2}\right) + \left(\frac{125}{60} - \frac{27}{60}\right)$

$= 2.43 + 1.6333\ldots = \frac{1219}{300} = 4.0633\ldots$

$\text{Var}(T) = E(T^2) - [E(T)]^2 = 4.0633\ldots - 1.66^2$

$= 1.3077\ldots = 1.31$ (3 s.f.)

(b) $0 \leq t \leq 3$: $F(t) = \int_0^t \frac{1}{50}(18 - 2x)\,dx = \frac{1}{50}(18t - t^2)$

$F(3) = \frac{1}{50}(54 - 9) = 0.9$

$3 < t \leq 5$: $F(t) = 0.9 + \int_3^t \frac{1}{20}\,dx = 0.9 + \frac{t-3}{20} = \frac{t}{20} + 0.75$

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

Leave blank

**Question 5 continued**

$$F(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{50}(18t - t^2) & 0 \leqslant t \leqslant 3 \\ \frac{t}{20} + 0.75 & 3 < t \leqslant 5 \\ 1 & t > 5 \end{cases}$$

(c) $P(T > 2) = 1 - F(2) = 1 - \frac{1}{50}(18 \times 2 - 2^2)$

$= 1 - \frac{32}{50} = \frac{9}{25} = 0.36$

(d) $E(T) - 2 = -0.34, \quad E(T) + 2 = 3.66$

$P(-0.34 < T < 3.66) = F(3.66) - F(-0.34) = F(3.66) - 0$

$= \frac{3.66}{20} + 0.75$

$= 0.933$

Leave blank

**Question 5 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

Leave blank

**Question 5 continued**

**(Total 13 marks)**

**Q5**

Leave blank

**6.** (a) Explain what you understand by the sampling distribution of a statistic.

**(1)**

A factory produces beads in bags for craft shops. A small bag contains 40 beads, a medium bag contains 80 beads and a large bag contains 150 beads. The factory produces small, medium and large bags in the ratio 5:3:2 respectively.

A random sample of 3 bags is taken from the factory.

(b) Find the sampling distribution for the range of the number of beads in the 3 bags in the sample.

**(7)**

A random sample of $n$ sets of 3 bags is taken. The random variable $Y$ represents the number of these $n$ sets of 3 bags that have a range of 70

(c) Calculate the minimum value of $n$ such that $P(Y = 0) < 0.2$

**(3)**

(a) The sampling distribution is all the possible values of the statistic together with their associated probabilities (i.e. the probability distribution of the statistic).

(b) $P(40) = 0.5$, $P(80) = 0.3$, $P(150) = 0.2$

Possible values of the range $R$ : 0, 40, 70, 110

$R = 0$ : (40,40,40), (80,80,80), (150,150,150)

$P(R = 0) = 0.5^3 + 0.3^3 + 0.2^3 = 0.16$

$R = 40$ : (40,40,80), (40,80,80)

$P(R = 40) = 3(0.5^2 \times 0.3) + 3(0.5 \times 0.3^2) = 0.225 + 0.135 = 0.36$

$R = 70$ : (80,80,150), (80,150,150)

$P(R = 70) = 3(0.3^2 \times 0.2) + 3(0.3 \times 0.2^2) = 0.054 + 0.036 = 0.09$

$R = 110$ : (40,40,150), (40,150,150), (40,80,150)

$P(R = 110) = 3(0.5^2 \times 0.2) + 3(0.5 \times 0.2^2) + 6(0.5 \times 0.3 \times 0.2)$

$= 0.15 + 0.06 + 0.18 = 0.39$

| $r$ | 0 | 40 | 70 | 110 |
|---|---|---|---|---|
| $P(R = r)$ | 0.16 | 0.36 | 0.09 | 0.39 |

(c) $Y \sim B(n, 0.09)$

$P(Y = 0) = 0.91^n < 0.2$

$n > \frac{\log 0.2}{\log 0.91} = 17.065...$

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

Leave blank

**Question 6 continued**

$(0.91^{17} = 0.2012\ldots, \quad 0.91^{18} = 0.1831\ldots < 0.2)$

minimum $n = 18$

Leave blank

**Question 6 continued**

Q6

**(Total 11 marks)**

**TOTAL FOR PAPER: 75 MARKS**

**END**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA