Please check the examination details below before entering your candidate information

Candidate surname	Other names
Annotated by Tam	

Centre Number | Candidate Number

HastingsMath.com

Pearson Edexcel International Advanced Level

Time 1 hour 30 minutes

Paper reference **WST01/01**

Mathematics

International Advanced Subsidiary/Advanced Level

Statistics S1

June 2022

You must have:
Mathematical Formulae and Statistical Tables (Yellow), 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 ▶

P71200A

Q:1/1/1/1/

P 7 1 2 0 0 A 0 1 2 4

Pearson

1. The company *Seafield* requires contractors to record the number of hours they work each week. A random sample of 38 weeks is taken and the number of hours worked per week by contractor Kiana is summarised in the stem and leaf diagram below.

Key : 3|2 means 32

Stem	Leaf											
1	4	4	4	5	5	5	6	6	9	9	9	(11)
2	1	2	2	3	3	4	4	4	w	9		(10)
3	2	3	4	4	5	6	7	7	7	9		(10)
4	1	1	2	3								(4)
5	1	9										(2)
6	4											(1)

The quartiles for this distribution are summarised in the table below.

Q_1	Q_2	Q_3
x	26	y

(a) Find the values of w, x and y

(3)

Kiana is looking for outliers in the data. She decides to classify as outliers any observations greater than

$$Q_3 + 1.0 \times (Q_3 - Q_1)$$

(b) Showing your working clearly, identify any outliers that Kiana finds.

(2)

(c) Draw a box plot for these data in the space provided on the grid opposite.

(3)

(d) Use the formula

$$\text{skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

to find the skewness of these data. Give your answer to 2 significant figures.

(2)

Kiana's new employer, *Landacre*, wishes to know the average number of hours per week she worked during her employment at *Seafield* to help calculate the cost of employing her.

(e) Explain why *Landacre* might prefer to know Kiana's mean, rather than median, number of hours worked per week.

(1)

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

Question 1 continued

(a) W: Q_2 position $= \frac{n}{2} = \frac{38}{2} = 19$

but even so 19;5 ∴ average 19th 20th

24 20+4

$Q_2 = 26$ (given)

∴

average: $\frac{24 + 20 + w}{2} = 26$

∴ $w = 8$ ✓

x: position Q1 $\frac{n}{4}$ $= 9{,}5^{th}$ term (round up)

10th term

$x = Q_1 = 19$ ✓

y: position Q3 $\frac{3n}{4} = \frac{3(38)}{4} = 28.5^{th}$ (round up)

29th term

$y = Q_3 = 37$ ✓

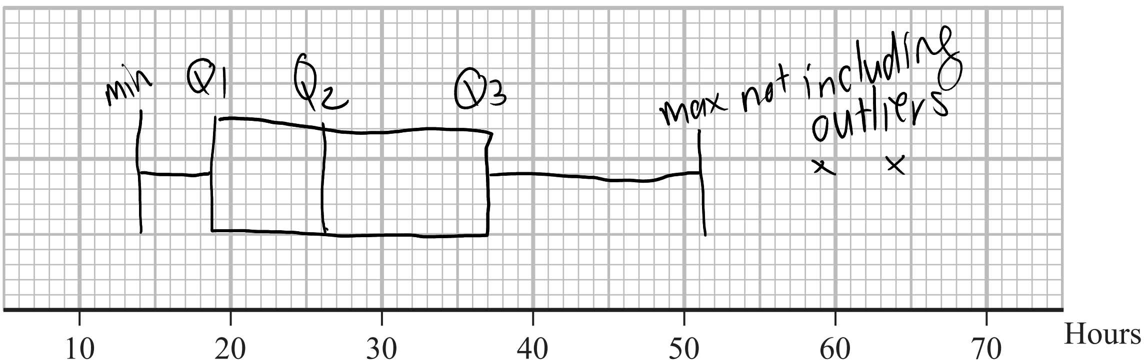
(b) $Q_3 = 37$ $Q_1 = 19$

∴ $37 + 1(37 - 19) = 55$ for UL

terms greater than 55

∴ 59; 64

(c)



Turn over for a spare grid if you need to redraw your box plot.

Question 1 continued

d) $\dfrac{(Q_3 - Q_2) - (Q_2 - Q)}{(Q_3 - Q_1)}$ = skewness

$Q_1 = 19$
$Q_2 = 26$
$Q_3 = 37$

sub in values

$\therefore$ skewness $= \dfrac{2}{9} = 0.22$

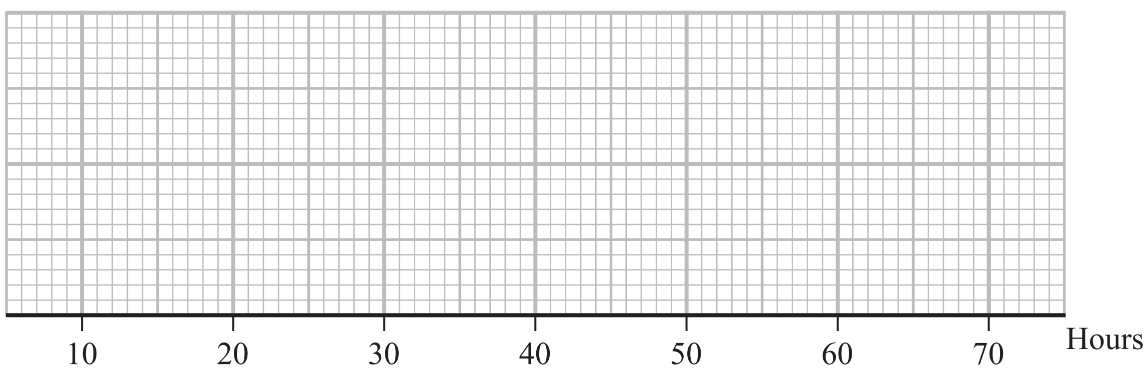
↳ positively skewed

e) Mean uses all data values
this reflects actual cost better.

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

Question 1 continued

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



(Total for Question 1 is 11 marks)

2. Stuart is investigating the relationship between Gross Domestic Product (GDP) and the size of the population for a particular country.
He takes a random sample of 9 years and records the size of the population, t millions, and the GDP, g billion dollars for each of these years.

The data are summarised as

$$n = 9 \qquad \sum t = 7.87 \qquad \sum g = 144.84 \qquad \sum g^2 = 3624.41 \qquad S_{tt} = 1.29 \qquad S_{tg} = 40.25$$

(a) Calculate the product moment correlation coefficient between t and g **(3)**

(b) Give an interpretation of your product moment correlation coefficient. **(1)**

(c) Find the equation of the least squares regression line of g on t in the form $g = a + bt$ **(4)**

(d) Give an interpretation of the value of b in your regression line. **(1)**

(e) (i) Use the regression line from part (c) to estimate the GDP, in billions of dollars, for a population of 7 000 000 **(2)**

(ii) Comment on the reliability of your answer in part (i). Give a reason, in context, for your answer. **(1)**

Using the regression line from part (c), Stuart estimates that for a population increase of x million there will be an increase of 0.1 billion dollars in GDP.

(f) Find the value of x **(2)**

(a) $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ $\quad \therefore r = \dfrac{S_{tg}}{\sqrt{S_{tt} \times S_{gg}}}$

$= \dfrac{40.25}{\sqrt{1.29 \times S_{gg}}}$

$S_{gg} = \Sigma g^2 - \dfrac{(\Sigma g)^2}{n}$

$= 3624.41 - \dfrac{144.84^2}{9} = 1293.4516$

Question 2 continued

$\therefore r = \dfrac{40.25}{\sqrt{1.29 \times 1293.4516}} = 0.985$

(b) Strong positive correlation.
As population increases; GDP increases

(c) $g = a + bt$

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

$b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{S_{tg}}{S_{tt}} = \dfrac{40.25}{1.29} = \dfrac{4025}{129}$

$= 31.2$ (3 sf)

$a = \dfrac{144.84}{9} - \dfrac{4025}{129}\left(\dfrac{7.87}{9}\right)$

$= -11.2$ (3 sf)

$\therefore g = -11.2 + 31.2t$

(d) Interpretation g : GDP t : population

use gradient : $31.2 = m = \dfrac{\Delta g}{\Delta t} = \dfrac{31.2}{1}$

every 1 mill population increase; GDP increases by 31.2 billion

(e)(i) $t = 7$

$g = -11.2 + 31.2(7)$

$= 237.2$ billion.

(ii) Reliability Unreliable because:

one piece of data can't be 7 000 000 → much larger than any t-values would ever be

$\bar{t} = \dfrac{7.87}{9} = 0.874$

Question 2 continued

g: GDP   t: population

f) $\frac{\Delta g}{\Delta t} = \frac{0.1}{x} = 31.2 \rightarrow$ gradient of least square regression line

gradient of least square regression line

$x = 0.00321$ (3sf)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 14 marks)

3. Gill buys a bag of logs to use in her stove. The lengths, l cm, of the 88 logs in the bag are summarised in the table below.

Length (l)	**Frequency (f)**
$15 < l \leqslant 20$	19
$20 < l \leqslant 25$	35
c $25 \overset{26}{<} l \leqslant 27$	16 $\frac{16}{2}$ for 26 (can assume)
$27 < l \leqslant 30$	15
$30 < l \leqslant 40$	3

A histogram is drawn to represent these data.
The bar representing logs with length $27 < l \leqslant 30$ has a width of 1.5 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing log lengths of $20 < l \leqslant 25$ **(3)**

(b) Use linear interpolation to estimate the median of l **(2)**

up to 26 in class (l)

The maximum length of log Gill can use in her stove is 26 cm.
Gill estimates, using linear interpolation, that x logs from the bag will fit into her stove.

(c) Show that $x = 62$ **(1)**

Gill randomly selects 4 logs from the bag. P(All four fit)

(d) Using $x = 62$, find the probability that all 4 logs will fit into her stove. **(2)**

Lo one's that fit

The weights, W grams, of the logs in the bag are coded using $y = 0.5w - 255$ and summarised by

$$n = 88 \qquad \sum y = 924 \qquad \sum y^2 = 12\,862$$

(e) Calculate

(i) the mean of W **(3)**

(ii) the variance of W **(3)**

DO NOT WRITE IN THIS AREA

Question 3 continued

width = ? height = ?

(a) $20 < l \leq 25$

$W = 25 - 20 = 5$ units

$27 < l \leq 30$ has width 1.5cm | height of 4cm (given)

$\therefore$ W: $1.5cm = 30 - 27$
$= 3$ units

$\therefore$ 1.5cm is 3 units

for 5 units: $\frac{1.5}{3} : \frac{3}{3}$

$\frac{1.5}{3} = 1$ unit $\therefore \frac{1.5}{3} \times 5 = 5$ units

$20 < l \leq 25$: $\therefore W = 2.5$ cm ✓

→ height is determined by fr. density

fr density = $\frac{\text{fr.}}{\text{interval}}$

$20 < l < 25$ fr density $= 35 \div 5$
$= 7$ $\therefore$ height = 7 units of fr. density

given info
$20 < l \leq 30$ frequency density $= 15 \div 3$
$= 5$ $\therefore$ height = 5 units of fr density

$\therefore$ 4cm (given) = 5 units of fr. density

$\frac{4cm}{5} = \frac{5 \text{ units}}{5}$

1 unit $= \frac{4}{5}$

$20 < l < 30$

$\therefore 7 \times \frac{4}{5} = 5.6$ cm = height ✓

Question 3 continued

(b) interpolation.

Q_2 = median position $\frac{n}{2} = \frac{88}{2} = 44$

class: $20 < l < 25$

$Q_2 \rightarrow 20 + \frac{44 - 19}{35} \times 5$ [or own interpolation method]

$Q_2 = 23.5714 = 23.6$ (3sf)

(c) $19 + 35 + \frac{16}{2} \rightarrow$ ends on 26 ∴ half way through fr.

$\downarrow$ fr

$= 62\text{cm}$ ✓

(d) $x = 62$ logs fit into the stove

$P(\text{one log}) = \frac{62}{88}$ $62 \rightarrow$ logs that fit, $88 \rightarrow$ total logs

$P(4 \text{ logs}) = \frac{62}{88} \times \frac{61}{87} \times \frac{60}{86} \times \frac{59}{85} = 0.239$ (3sf)

∴ 0.239 is the probability that all four of the randomly selected logs will fit in the stove

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

(e) (i) weights w grams

$y = 0.5w - 255$ coding

mean of w $\bar{w} = ?$ (not coded)

$n = 88$

coded: $\Sigma y = 924$

coded: $\Sigma y^2 = 12862$

↳ mean of y: $\bar{y} = \frac{\Sigma y}{n} = \frac{924}{88}$

$= \frac{21}{2} = 10.5$ ✓

$\bar{y} = 0.5\bar{w} - 255$ (mean is affected in coding by ± and × ÷)

$\frac{\bar{y} + 255}{0.5} = \bar{w} =$

$= 531$ grams ✓

(ii) $Var(W) = ?$

$Var(y) = \frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2$ (ms − sm)

$= \frac{12862}{88} - \left(\frac{924}{88}\right)^2$

$= \frac{395}{11}$

$Var(W) = ?$ (+ − does not affect variance) (only × ÷)

$Var(Y) = (0.5)^2 \times Var(W)$

$\frac{395}{11} = (0.5)^2 \times Var(W)$

(Total for Question 3 is 14 marks)

$Var(W) = \frac{395}{11} \div (0.5)^2 = 144$ (3sf)

4. The events H and W are such that

$$P(H) = \frac{3}{8} \qquad P(H \cup W) = \frac{3}{4}$$

Given that H and W are independent, $P(H \cap W) = P(H) \times P(W)$

(a) show that $P(W) = \frac{3}{5}$

(4)

The event N is such that

$$P(N) = \frac{1}{15} \qquad P(H \cap N) = P(N)$$

(b) Find $P(N' | H)$

(2)

Given that W and N are mutually exclusive,

(c) draw a Venn diagram to represent the events H, W and N giving the exact probabilities of each region in the Venn diagram.

(5)

a $P(H \cap W) = P(H) \times P(W)$ (independent)

$P(H) = \frac{3}{8}$ (given)

$P(H \cup W) = \frac{3}{4}$ (given)

$P(H \cup W) = P(H) + P(W) - P(H \cap W)$

$P(H \cap W) = \frac{3}{8} \times P(W)$

$\therefore \frac{3}{4} = \frac{3}{8} + P(W) - \frac{3}{8} P(W)$

$6 = 3 + 8P(W) - 3P(W)$

$6 = 5P(W)$

$\frac{3}{5} = P(W)$ ✓

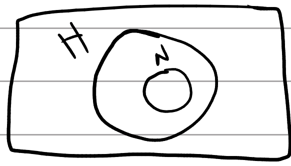
DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 4 continued

(b) $P(H) = \frac{1}{8}$ (given)

$P(N) = \frac{1}{15}$ (given)

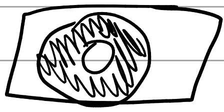
$P(H \cap N) = P(N)$ (given)



$P(N'|H) = ?$

$P(N'|H) = \frac{P(N' \cap H)}{P(H)}$

$P(N' \cap H)$



visually

$= P(H) - P(N)$ (by inspec)

$= \frac{3}{8} - \frac{1}{15} = \frac{37}{120}$

$P(N'|H) = \frac{\frac{37}{120}}{\frac{3}{8}} = \frac{37}{48}$

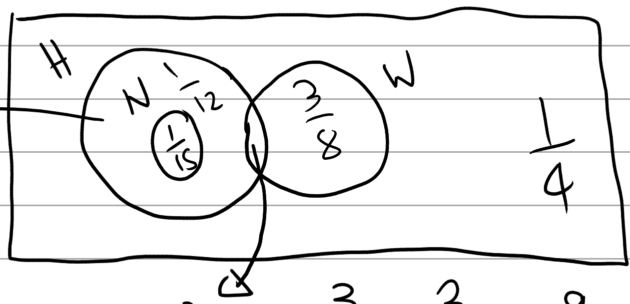
(c) $P(H) = \frac{3}{8}$ $P(H \cup W) = \frac{3}{4}$ $P(W) = \frac{3}{5}$ $P(N' \cap H) = \frac{37}{120}$

$P(N) = \frac{1}{15}$ $P(H \cap N) = P(N)$ $P(N'|H) = \frac{37}{48}$

$P(H\text{only})$

$= \frac{3}{8} - \frac{1}{15} - \frac{9}{45}$

$= \frac{1}{12}$



$P(H \cap W) = P(H) \times P(W) = \frac{3}{8} \times \frac{3}{5} = \frac{9}{14}$

$P(W\text{only}) = \frac{3}{5} - \frac{9}{40} = \frac{3}{8}$

Question 4 continued

Question 4 continued

(Total for Question 4 is 11 marks)

discrete random variables

5. A red spinner is designed so that the score R is given by the following probability distribution.

(e) Pabel

r	2	3	4	5	6
$\mathrm{P}(R = r)$	0.25	0.3	0.15	0.1	0.2

(a) Show that $\mathrm{E}(R^2) = 15.8$ **(1)**

Given also that $\mathrm{E}(R) = 3.7$

(b) find the standard deviation of R, giving your answer to 2 decimal places. **(2)**

A yellow spinner is designed so that the score Y is given by the probability distribution in the table below. The cumulative distribution function $\mathrm{F}(y)$ is also given.

r^2 4 9 16 25 36

(e) Jessie

y	2	3	4	5	6
$\mathrm{P}(Y = y)$	0.1	0.2	0.1	0.25 a	0.35 b
cum fr = $\mathrm{F}(y)$	0.1	0.3	0.4	0.65 c	d 1

$c = 0.4 + 0.25$

(c) Write down the value of d **(1)**

Given that $\mathrm{E}(Y) = 4.55$

(d) find the value of c **(5)**

Pabel and Jessie play a game with these two spinners.
Pabel uses the red spinner.
Jessie uses the yellow spinner.
They take turns to spin their spinner.
The winner is the first person whose spinner lands on the number 2 and the game ends.
Jessie spins her spinner first.

(e) Find the probability that Jessie wins on her second spin. **(2)**

(f) Calculate the probability that, in a game, the score on Pabel's first spin is the same as the score on Jessie's first spin. **(3)**

DO NOT WRITE IN THIS AREA

Question 5 continued

(a) $E(r^2) = 4\times0.25 + 9\times0.3 + \dots + 36\times0.2$
$= 15.8$

(b) $E(r) = 3.7$
$S.D = ?$
$S.D = \sqrt{E(R)^2 - [E(R)]^2}$
$= \sqrt{15.8 - 3.7^2}$
$S.D = \frac{\sqrt{211}}{10}$
$= 1.45258$
$= 1.45$ ✓

(c) $d = 1$ (cum fr)

(d) $E(Y) = 2\times0.1 + 3\times0.2 + \dots + 5\times a + 6\times b$ ~~6×b~~
$4.55 =$
$4.55 = 5a + 6b + 1.2$
$5a + 6b = 3.35$ .... (1)

$\Sigma p(Y=y) = 1$
$\therefore 0.1 + 0.2 + \dots + a + b = 1$
$a + b = 0.6$
$b = 0.6 - a$ .. (2)

Sub (2) into (1)

Question 5 continued

$5a + 6(0.6 - a) = 3.35$
$5a + 3.6 - 6a = 3.35$
$-a = -0.25$
$a = 0.25$
$\therefore b = 0.6 - 0.25$
$= 0.35$

$c = 0.4 + 0.25$ (cum fr)
$= 0.65$ ✓

(e) $\boxed{\begin{array}{l} P : RS \\ J : YS \end{array}}$

Jessie spins first

THROWS: (1) $P(J \neq 2)$ × (2) $P(P \neq 2)$ × (3) $P(J = 2)$

(1) $P(J \neq 2) = 1 - (P(J=2))$
$= 1 - 0.1$
$= 0.9$ and

(2) $P(P \neq 2) = 1 - (P(J=2))$
$= 1 - 0.25$
$= 0.75$ and

(3) $P(J \neq 2) = 0.1$

$\therefore 0.9 \times 0.75 \times 0.1 = 0.0675$ ✓

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 5 continued

(f) $P(22)$ or $+$ $P(33)$ or $+ P(44)$ or $+ P(55)$ or $+ P(66)$

$\therefore P(J=2) \times P(P=2) + P(J=3) \times P(P=3) + \dots + P(J=6) \times P(P=6)$

$\therefore 0.1 \times 0.25 + 0.2 \times 0.3 + 0.1 \times 0.15 + 0.25 \times 0.1 + 0.35 \times 0.2$

$= 0.195$ ✓

but exclude $P(22)$ because if J or P gets a 2 the game stops

$\therefore 0.2 \times 0.3 + \dots + 0.35 + 0.2$

$= 0.17$ ✓

(Total for Question 5 is 14 marks)

→ normal distribution

μ σ^2

6. A manufacturer fills bottles with oil.
The volume of oil in a bottle, V ml, is normally distributed with $V \sim N(100, 2.5^2)$

(a) Find $P(V > 104.9)$

$\mu = 100$ $SD = 2.5$

(3)

(b) In a pack of 150 bottles, find the expected number of bottles containing more than 104.9 ml

(2)

→ positive value

(c) Find the value of v, to 2 decimal places, such that $P(V > v \mid V < 104.9) = 0.2801$

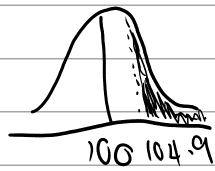
(6)

a $P(V > 104.9)$

$= P\left(Z > \frac{104.9 - 100}{2.5}\right)$

$= P(Z > 1.96)$

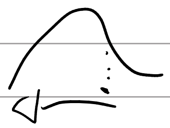
V:



Z: 0 1.96

$Z = \frac{x - \mu}{\sigma}$

go to tables



$1 - P(Z < 1.96)$

$\therefore 1 - 0.9750$ → from tables

$= 0.025$ ✓

b 150 bottles

$E(X) = np$

$= 150 \times 0.025$ → from a

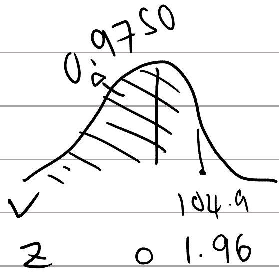
$= 3.75$ bottles ✓

Question 6 continued

(c) $v = ?$ 2 decimal places $\longrightarrow$ conditional probability.

$$P(V > v \mid V < 104.9) = 0.2801$$

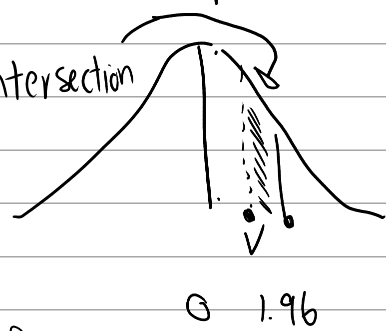
$$= \frac{P(V > v \cap V < 104.9)}{P(V < 104.9)} = 0.2801$$



$$= \frac{}{0.9750}$$

but v positive

$\therefore$ intersection

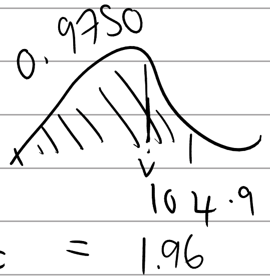


$$= \frac{P(V > v) \cap P(V < 104.9)}{0.9750} = 0.2801$$

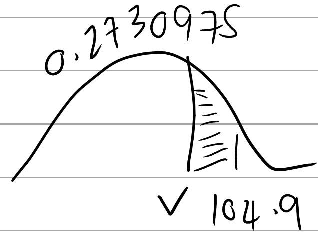
$$= P(V > v) \cap P(V < 104.9) = 0.2801 \times 0.9750$$

$$= 0.2730975$$

$\therefore$



$-$



$$= 0.9750 - 0.2750975$$

tables

$$= 0.7019025 \quad \text{(Area)}$$

$\therefore z = 0.53 \longrightarrow$ from tables

v: $z = \dfrac{v - \mu}{\sigma}$ $\quad V \sim N(100, 2.5^2)$

Question 6 continued

$$0.53 = \frac{V - 100}{2.5}$$

$$V = 0.53 \times 2.5 + 100$$
$$= 101.325$$

$\therefore$ 101.33 (2 dec places) ✓

(Total for Question 6 is 11 marks)

TOTAL FOR PAPER: 75 MARKS

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