PhysicsAndMathsTutor.com

Write your name here

Surname	Other names
MODEL ANSWERS	Chris

Pearson Edexcel GCE

Centre Number | Candidate Number

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 15 June 2016 – Morning
Time: 1 hour 30 minutes

Paper Reference
6683/01

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

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

P46673A

©2016 Pearson Education Ltd.

1/1/1/1/

P 4 6 6 7 3 A 0 1 2 4

PEARSON

Leave blank

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

1. A biologist is studying the behaviour of bees in a hive. Once a bee has located a source of food, it returns to the hive and performs a dance to indicate to the other bees how far away the source of the food is. The dance consists of a series of wiggles. The biologist records the distance, d metres, of the food source from the hive and the average number of wiggles, w, in the dance.

Distance, d m	30	50	80	100	150	400	500	650
Average number of wiggles, w	0.725	1.210	1.775	2.250	3.518	6.382	8.185	9.555

[You may use $\sum w = 33.6$ $\sum dw = 13\,833$ $S_{dd} = 394\,600$ $S_{ww} = 80.481$ (to 3 decimal places)]

(a) Show that $S_{dw} = 5601$
(2)

(b) State, giving a reason, which is the response variable.
(1)

(c) Calculate the product moment correlation coefficient for these data.
(2)

(d) Calculate the equation of the regression line of w on d, giving your answer in the form $w = a + bd$
(4)

A new source of food is located 350 m from the hive.

(e) (i) Use your regression equation to estimate the average number of wiggles in the corresponding dance.

(ii) Comment, giving a reason, on the reliability of your estimate.
(2)

a) $\sum d = 30 + 50 + 80 + 100 + 150 + 400 + 500 + 650$
$= 1960$

$S_{dw} = \sum dw - \frac{\sum w \sum d}{n}$

$= 13\,833 - \frac{33.6 \times 1960}{8}$

$S_{dw} = 5601$

✓

P 4 6 6 7 3 A 0 2 2 4

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

Leave blank

Question 1 continued

b) Distance, d, is the independent variable, what is changed in the study.

Number of wiggles is the dependent variable, what is measured in the study.

c) Product Moment Correlation Coefficient

$$r = \frac{S_{dw}}{\sqrt{S_{dd} \; S_{ww}}}$$

$$= \frac{5601}{\sqrt{394600 \times 80.481}}$$

$r = 0.99389 \ldots$

$r \simeq 0.994$ (3 D.P.)

d) Regression Line

$$b = \frac{S_{dw}}{S_{dd}} = \frac{5601}{394600} = 0.014194\ldots$$

$b \simeq 0.0142$ (3 sig fig)

$$\bar{d} = \frac{\Sigma d}{n} = \frac{1960}{8} = 245$$

$$\bar{w} = \frac{\Sigma w}{n} = \frac{33.6}{8} = 4.2$$

Leave blank

Question 1 continued

Regression line in the form

$y = a + bx$

So $w = a + bd$

Where $a = \bar{w} - b\bar{d}$

$a = 4.2 - 0.0142 \times 245$

$a = 0.72244\ldots \simeq \underline{0.722}$ (3 sig fig)

$\therefore \underline{w = 0.722 + 0.0142 d}$

e)(i) If $d = 350$m

$w = 0.722 + 0.0142 (350)$

$w = 5.69034\ldots$

$\underline{w \simeq 5.69}$ (3 sig. fig.)

(ii) As the PMCC is close to 1 there is a strong positive correlation between the two variables, w and d.

As 350m is within the range of the data, the estimate is based on interpolation and is reliable.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

P 4 6 6 7 3 A 0 4 2 4

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 11 marks)

Leave blank

2. The discrete random variable X has the following probability distribution, where p and q are constants.

x	-2	-1	$\frac{1}{2}$	$\frac{3}{2}$	2
$\mathrm{P}(X = x)$	p	q	0.2	0.3	p

(a) Write down an equation in p and q

(1)

Given that $\mathrm{E}(X) = 0.4$

(b) find the value of q

(3)

(c) hence find the value of p

(2)

Given also that $\mathrm{E}(X^2) = 2.275$

(d) find $\mathrm{Var}(X)$

(2)

Sarah and Rebecca play a game.

A computer selects a single value of X using the probability distribution above.

Sarah's score is given by the random variable $S = X$ and Rebecca's score is given by the random variable $R = \dfrac{1}{X}$

(e) Find $\mathrm{E}(R)$

(3)

Sarah and Rebecca work out their scores and the person with the higher score is the winner. If the scores are the same, the game is a draw.

(f) Find the probability that

(i) Sarah is the winner,

(ii) Rebecca is the winner.

(4)

a) $\sum P(X=x) = p + q + 0.2 + 0.3 + p = 1$

$\underline{2p + q = 0.5}$ ✓

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

P 4 6 6 7 3 A 0 6 2 4

DO NOT WRITE IN THIS AREA

Leave blank

Question 2 continued

b) $E(X) = \sum x\, P(X=x)$

$= \cancel{-2p} - 1q + \frac{1}{2}(0.2) + \frac{3}{2}(0.3) + \cancel{2p}$

$0.4 = -q + 0.1 + 0.45$

$\underline{q = 0.15}$

c) From (a)

$2p + q = 0.5$

$2p = 0.5 - (0.15)$

$p = 0.35/2$

$\underline{p = 0.175}$

d) $Var(X) = \sum x^2 P(X=x) - (E(X))^2$

$\sum x^2 P(X=x) = (-2)^2 p + (-1)^2 q + (\frac{1}{2})^2 0.2 + (\frac{3}{2})^2 0.3 + (2)^2 p$

$= 4\times0.175 + 1\times0.15 + \frac{1}{4}\times0.2 + \frac{9}{4}\times0.3 + 4\times0.175$

$= 0.7 + 0.15 + 0.05 + \frac{27}{40} + 0.7$

$= 1.4 + 0.2 + 0.675$

$\underline{= 2.275}$

0.675
$4\overline{)2.2700}$

$\therefore Var(X) = 2.275 - 0.4^2$

$0.4^2 = 0.16$

$\underline{= 2.115}$

Turn over

P46673A0724

Leave blank

Question 2 continued

$r=\frac{1}{x}$	$\frac{-1}{2}$	$\frac{-1}{1}$	$\frac{2}{1}$	$\frac{2}{3}$	$\frac{1}{2}$
$P(R=r)$	0.175	0.15	0.2	0.3	0.175

$E(R) = \sum r\,P(X=x)$

$= \cancel{-\frac{1}{2}\times 0.175} - 1\times 0.15 + 2\times 0.2 + \frac{2}{3}\times 0.3 \cancel{+\frac{1}{2}\times 0.175}$

$= -0.15 + 0.4 + 0.2$

$E(R) = 0.45$

f) Sarah > Rebecca, if $x = \frac{3}{2}$, $x = 2$

$P(\text{Sarah wins}) = 0.3 + 0.175 = 0.475$

Rebecca > Sarah, if $x = -2$, $x = \frac{1}{2}$

$P(\text{Rebecca wins}) = 0.175 + 0.2 = 0.375$

DO NOT WRITE IN THIS AREA

P46673A0824

Leave blank

Question 2 continued

b) $S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{\Sigma f}$ (where $\Sigma f = n$)

$= \Sigma f \ \sigma_y^2$

$= 8 \times 9.461^2$

$= 716.084...$

$S_{yy} \simeq 716$ (3 sig fig)

c) $S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{\Sigma f}$

$= 4900.5 - \frac{86.8 \times \bar{y} \times \Sigma f}{8}$ *

$= 4900.5 - \frac{86.8}{\cancel{8}} \times 58 \times \cancel{8}$

$S_{xy} = -133.9$

* Because $\bar{y} = \frac{\Sigma y}{\Sigma f}$

Q2

(Total 15 marks)

P46673A0924

Turn over

Leave blank

3. Before going on holiday to *Seapron*, Tania records the weekly rainfall (x mm) at *Seapron* for 8 weeks during the summer. Her results are summarised as

$$\sum x = 86.8 \qquad \sum x^2 = 985.88$$

(a) Find the standard deviation, σ_x, for these data.
(3)

Tania also records the number of hours of sunshine (y hours) per week at *Seapron* for these 8 weeks and obtains the following

$$\bar{y} = 58 \qquad \sigma_y = 9.461 \text{ (correct to 4 significant figures)} \qquad \sum xy = 4900.5$$

(b) Show that $S_{yy} = 716$ (correct to 3 significant figures)
(1)

(c) Find S_{xy}
(2)

(d) Calculate the product moment correlation coefficient, r, for these data.
(2)

During Tania's week-long holiday at *Seapron* there are 14 mm of rain and 70 hours of sunshine.

(e) State, giving a reason, what the effect of adding this information to the above data would be on the value of the product moment correlation coefficient.
(2)

a) Variance $\sigma_x^2 = \dfrac{\sum x^2}{\sum f} - \left(\dfrac{\sum x}{\sum f}\right)^2$

$\sigma_x^2 = \dfrac{985.88}{8} - \left(\dfrac{86.8}{8}\right)^2$

$\sigma_x^2 = \dfrac{441}{80}$

Standard deviation $\sigma_x = \sqrt{\dfrac{441}{80}} = \dfrac{21\sqrt{5}}{20}$

$\sigma_x = 2.34787\ldots \simeq 2.35$

✓

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

P 4 6 6 7 3 A 0 1 0 2 4

DO NOT WRITE IN THIS AREA

Leave blank

Question 3 continued

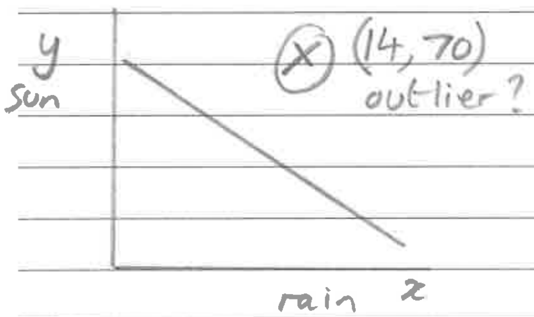
d) PMCC :

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \quad S_{yy}}}$$

$$= \frac{-133.9}{\sqrt{\frac{441}{10} \times 716}}$$

$$= {}^{-}0.75349\ldots$$

$$r \simeq {}^{-}0.753$$



Where $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{\Sigma f}$

$= 985.88 - \frac{86.8^2}{8}$

$S_{xx} = \frac{441}{10}$

e) Rain : $\bar{x} = \frac{86.8}{8} = 10.85$ mm

Adding 14mm rain would increase the mean
increase σ_x
so r decreases.

Sunshine : $\bar{y} = 58$ hours.

Adding 70 hours sunshine would increase the mean
increase σ_y
so r decreases.

Leave blank

Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

P 4 6 6 7 3 A 0 1 2 2 4

Leave blank

Question 3 continued

Q3

(Total 10 marks)

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

Leave blank

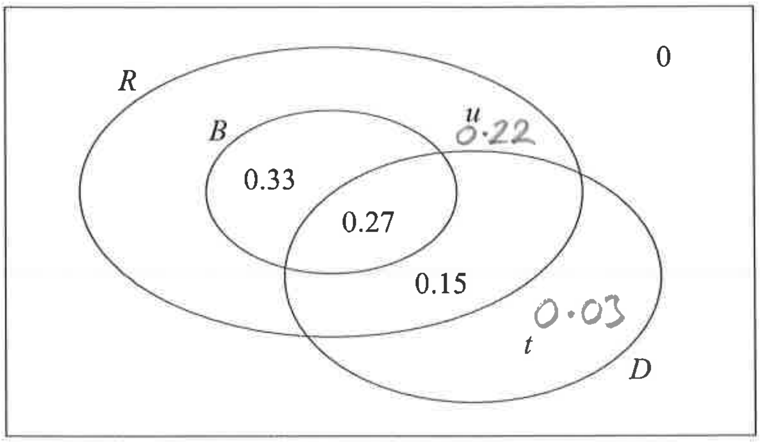
4. The Venn diagram shows the probabilities of customer bookings at Harry's hotel.

R is the event that a customer books a room

B is the event that a customer books breakfast

D is the event that a customer books dinner

u and t are probabilities.



S=1

(a) Write down the probability that a customer books breakfast but does not book a room. **(1)**

Given that the events B and D are independent

(b) find the value of t **(4)**

(c) hence find the value of u **(2)**

(d) Find

(i) $P(D|R \cap B)$

(ii) $P(D|R \cap B')$ **(4)**

A coach load of 77 customers arrive at Harry's hotel.

Of these 77 customers

40 have booked a room and breakfast

37 have booked a room without breakfast

(e) Estimate how many of these 77 customers will book dinner. **(2)**

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

P 4 6 6 7 3 A 0 1 4 2 4

DO NOT WRITE IN THIS AREA

Leave blank

Question 4 continued

a) $P(B \mid R') = 0$

b) B and D independent :

$P(B) \times P(D) = P(B \cap D)$

$(0.33+0.27) \times P(D) = 0.27$

$P(D) = \frac{0.27}{0.60} = \frac{9}{20} = 0.45$

$\Rightarrow 0.27 + 0.15 + t = 0.45$

$t = 0.03$

c) Adding probabilities :

$0.33 + 0.27 + 0.15 + 0.03 + u = 1$

$u = 0.22$

d) (i) $P(D \mid R \cap B) = \frac{0.27}{0.33+0.27} = 0.45$

(ii) $P(D \mid R \cap B') = \frac{0.15}{0.15+0.22} = \frac{15}{37}$

Leave blank

Question 4 continued

e) Using answers to part(d)

$P(D \mid R \cap B) = 0.45 \Rightarrow 0.45 \times 40 = 18$

$P(D \mid R \cap B') = \frac{15}{37} \Rightarrow \frac{15}{37} \times 37 = 15$

Sum total booking dinner $\underline{\overline{33}}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

P 4 6 6 7 3 A 0 1 6 2 4

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

Leave blank

Question 4 continued

Q4

(Total 13 marks)

Leave blank

5. A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	Frequency (f)	Weight midpoint (x)	Cumulative frequency
$0 \leqslant w < 2$	1	1	1
$2 \leqslant w < 3$	8	2.5	9
$3 \leqslant w < 3.5$	17	3.25	26
$3.5 \leqslant w < 4$	17	3.75	43
$4 \leqslant w < 5$	7	4.5	50

[You may use $\sum fx^2 = 611.375$] 15 Total

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leqslant w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leqslant w < 3.5$ **(3)**

(b) Use linear interpolation to estimate the median weight of these babies. **(2)**

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies. **(3)**

Shyam decides to model the weights of babies born at the hospital, by the random variable W, where $W \sim \mathrm{N}(3.43, 0.65^2)$

(d) Find $\mathrm{P}(W < 3)$ **(3)**

(e) With reference to your answers to (b), (c)(i) and (d) comment on Shyam's decision. **(3)**

A newborn baby weighing 3.43 kg is born at the hospital.

(f) Without carrying out any further calculations, state, giving a reason, what effect the addition of this newborn baby to the sample would have on your estimate of the

(i) mean, No change as weight same as mean. ✓

(ii) standard deviation. Decreases as data concentrated around mean/centre. ✓ (3)

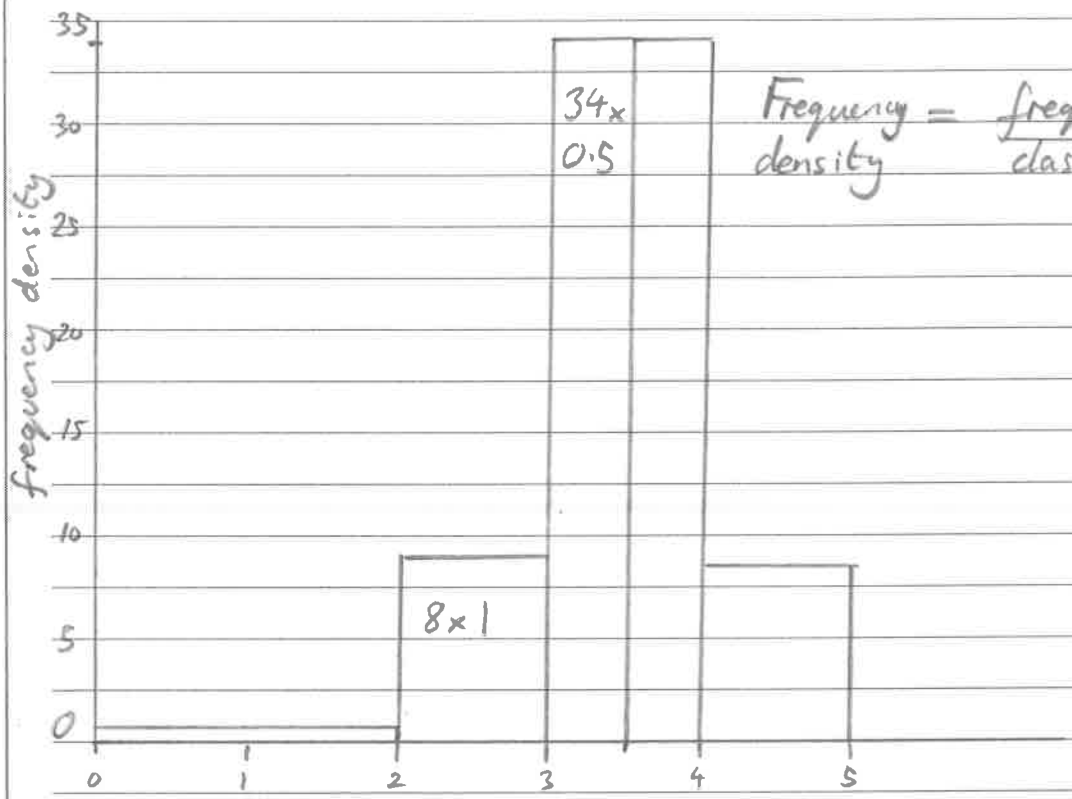
DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

P 4 6 6 7 3 A 0 1 8 2 4

Leave blank

Question 5 continued

Histogram



Frequency density = frequency / class width

a) If $2 \le w < 3$ $w = 1cm$ $h = 4cm$

Then $3 \le w < 3.5$ $w = 0.5cm$ $h = 4 \times \frac{34}{8} = 17cm$ ✓

b) Median : $\frac{50}{2}$ so 25^{th} term. Linear intopolation:

3.0	m	3.5
9	25	26

$$\frac{m - 3.0}{3.5 - 3.0} = \frac{25 - 9}{26 - 9}$$

Leave blank

Question 5 continued

$$m = \frac{16}{17} \times 0.5 + 3.0$$

$$m = \frac{59}{17} \simeq 3.47$$

c) Grouped mean:

$$\bar{x} = \frac{\sum fx}{\sum f}$$

$$= \frac{1\times1 + 8\times2.5 + 17\times3.25 + 17\times3.75 + 7\times4.5}{50}$$

$$\bar{x} = \frac{343}{100} = 3.43$$

Standard deviation:

$$\sigma_x = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$

$$= \sqrt{\frac{611.375}{50} - 3.43^2}$$

$$\sigma_x = 0.680$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

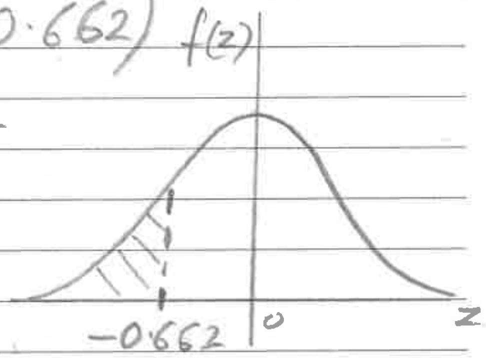
Leave blank

Question 5 continued

d) W represents weights of babies

$$W \sim N(3.43, 0.65^2)$$
$$(\mu, \sigma^2)$$

$$P(W<3) = P\left(Z < \frac{w-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{3-3.43}{0.65}\right)$$
$$= P(Z < -0.662)$$
$$= 1 - P(Z < 0.662)$$
$$= 1 - 0.7454$$
$$= 0.2546$$
$$\simeq 0.255$$



e) Median = 3.47

Mean = 3.43

The median is similar to the mean so the SKEW is low.

$P(W<3) = 0.255$ These are quite similar.

From histogram, number less than 3kg is $\frac{9}{50} = 0.18$.

(Total 17 marks)

Q5

Leave blank

6. The time, in minutes, taken by men to run a marathon is modelled by a normal distribution with mean 240 minutes and standard deviation 40 minutes.

(a) Find the proportion of men that take longer than 300 minutes to run a marathon.
(3)

Nathaniel is preparing to run a marathon. He aims to finish in the first 20% of male runners.

(b) Using the above model estimate the longest time that Nathaniel can take to run the marathon and achieve his aim.
(3)

The time, W minutes, taken by women to run a marathon is modelled by a normal distribution with mean μ minutes.

Given that $P(W < \mu + 30) = 0.82$

(c) find $P(W < \mu - 30 \mid W < \mu)$
(3)

a) $X \sim N(240, 40^2)$

(μ, σ^2)

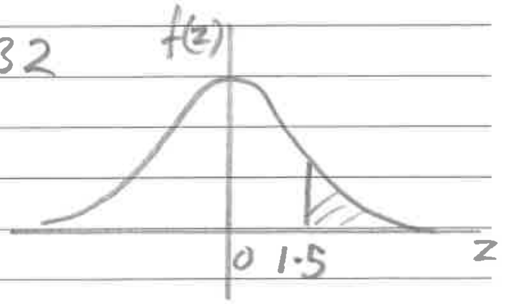
$P(X > 300) = P\left(Z > \frac{300-240}{40}\right)$

$= P(Z > 1.5)$

$= 1 - P(Z < 1.5)$

$= 1 - 0.9332$

$= \underline{0.0668}$



✓

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

Leave blank

Question 6 continued

b) $P(Z < z) = 0.2$

$z = 0.8416$ using % tables

$z = \frac{X - \mu}{\sigma}$

$\Rightarrow X = z\sigma + \mu$

$= 0.8416 \times 40 + 240$

$= 273.664$

$\simeq 274$ minutes

c) $W \sim N(\mu, \sigma^2)$

$P(W < \mu + 30) = P\left(Z < \frac{w - \mu}{\sigma}\right)$

$= P\left(Z < \frac{\mu + 30 - \mu}{\sigma}\right)$

$= P\left(Z < \frac{30}{\sigma}\right) = 0.82$

$\Rightarrow z = \frac{30}{\sigma} = 0.92$

$\sigma = \frac{30}{0.92} \simeq 32.6$

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

Leave blank

Question 6 continued

$$P(W < \mu - 30) = P\left(Z < \frac{\mu - 30 - \mu}{\sigma}\right)$$

$$= P\left(Z < \frac{-30}{\sigma}\right)$$

$$= 1 - P\left(Z < \frac{+30}{\sigma}\right)$$

$$= 1 - 0.82$$

$$= \underline{0.18}$$

$$P(W < \mu - 30 \mid W < \mu)$$

$$= \frac{0.18}{0.5} = \frac{9}{25}$$

$$= \underline{0.36}$$

Q6

(Total 9 marks)

TOTAL FOR PAPER: 75 MARKS

END

P 4 6 6 7 3 A 0 2 4 2 4

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