Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

Centre Number | Candidate Number

# Pearson Edexcel Level 3 GCE

## Monday 24 June 2024

Afternoon (Time: 1 hour 30 minutes)

Paper reference **9FM0/4B**

# Further Mathematics

## Advanced

## PAPER 4B: Further Statistics 2

**You must have:**
Mathematical Formulae and Statistical Tables (Green), 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 statistical tables should be quoted in full. If a calculator is used instead of 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 8 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.

*Turn over* ▶

P72801A

©2024 Pearson Education Ltd.
F:1/1/1/1/1/1/

P 7 2 8 0 1 A 0 1 2 8

Pearson

**1.** Two students are experimenting with some water in a plastic bottle. The bottle is filled with water and a hole is put in the bottom of the bottle. The students record the time, $t$ seconds, it takes for the water level to fall to each of 10 given values of the height, $h$ cm, above the hole.

Student $A$ models the data with an equation of the form $t = a + b\sqrt{h}$

The data is coded using $v = t - 40$ and $w = \sqrt{h}$ and the following information is obtained.

$$\sum v = 626 \quad \sum v^2 = 64678 \quad \sum w = 22.47 \quad S_{ww} = 4.52 \quad S_{vw} = -338.83$$

(a) Find the equation of the regression line of $t$ on $\sqrt{h}$ in the form $t = a + b\sqrt{h}$

**(4)**

The time it takes the water level to fall to a height of 9 cm above the hole is 47 seconds.

(b) Calculate the residual for this data point.
Give your answer to 2 decimal places.

**(2)**

Given that the residual sum of squares (RSS) for the model of $t$ on $\sqrt{h}$ is the same as the RSS for the model of $v$ on $w$,

(c) calculate the RSS for these 10 data points.

**(2)**

Student $B$ models the data with an equation of the form $t = c + dh$

The regression line of $t$ on $h$ is calculated and the residual sum of squares (RSS) is found to be 980 to 3 significant figures.

(d) With reference to part (c) state, giving a reason, whether Student $B$'s model or Student $A$'s model is the more suitable for these data.

**(1)**

(a) $b = \frac{S_{vw}}{S_{ww}} = \frac{-338.83}{4.52} = -74.9623...$

$\bar{v} = \frac{626}{10} = 62.6 \qquad \bar{w} = \frac{22.47}{10} = 2.247$

$a = \bar{v} - b\bar{w} = 62.6 - (-74.9623...)(2.247) = 231.0404...$

$v = 231.0404... - 74.9623...w$

$t - 40 = 231.0404... - 74.9623...\sqrt{h}$

$t = 271.04... - 74.96...\sqrt{h}$

$t = 271 - 75.0\sqrt{h}$

(b) $h = 9$: $\hat{t} = 271.0404... - 74.9623... \times \sqrt{9} = 46.1533...$

residual $= 47 - 46.1533... = 0.84667...$

residual $= 0.85$ (2 d.p.)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Question 1 continued**

(c) $S_{vv} = \Sigma v^2 - \frac{(\Sigma v)^2}{10} = 64678 - \frac{626^2}{10} = 25490.4$

$RSS = S_{vv} - \frac{(S_{vw})^2}{S_{ww}} = 25490.4 - \frac{(-338.83)^2}{4.52}$

$= 90.8936...$

$RSS = 90.9$

(d) Student A's model is more suitable, because the RSS for Student A's model (90.9) is much smaller than the RSS for Student B's model (980), so the data points lie closer to Student A's regression line.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**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

**Question 1 continued**

**(Total for Question 1 is 9 marks)**

**2.** An estate agent asks customers to rank 7 features of a house, *A*, *B*, *C*, *D*, *E*, *F* and *G*, in order of importance. The responses for two randomly selected customers are in the table below.

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Customer 1 | *A* | *E* | *C* | *F* | *G* | *B* | *D* |
| Customer 2 | *E* | *F* | *C* | *G* | *A* | *D* | *B* |

(a) Calculate Spearman's rank correlation coefficient for these data.

**(4)**

(b) Stating your hypotheses and critical value clearly, test at the 5% level of significance, whether or not the two customers are generally in agreement.

**(3)**

(a)

| Feature | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| Customer 1 rank | 1 | 6 | 3 | 7 | 2 | 4 | 5 |
| Customer 2 rank | 5 | 7 | 3 | 6 | 1 | 2 | 4 |
| $d$ | −4 | −1 | 0 | 1 | 1 | 2 | 1 |
| $d^2$ | 16 | 1 | 0 | 1 | 1 | 4 | 1 |

$\Sigma d^2 = 16 + 1 + 0 + 1 + 1 + 4 + 1 = 24$

$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 24}{7(7^2-1)} = 1 - \frac{144}{336}$

$= 0.571428... = \frac{4}{7}$

$r_s = 0.571$

(b) $H_0: \rho_s = 0$ $H_1: \rho_s > 0$

Critical value ($n = 7$, 5%, one-tailed): 0.7143

$0.571 < 0.7143$, so not significant: do not reject $H_0$

There is insufficient evidence to suggest that the two customers are generally in agreement (no evidence of positive correlation between their rankings of the features).

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 2 continued**

**(Total for Question 2 is 7 marks)**

**3.** A factory produces bolts. The lengths of the bolts are normally distributed with mean $\mu$ mm and standard deviation 0.868 mm

A random sample of 15 of these bolts is taken and the mean length is 30.03 mm

(a) Calculate a 90% confidence interval for $\mu$

**(3)**

A suitable test, at the 10% level of significance, is carried out using these 15 bolts, to see whether or not there is evidence that the variance of the length of the bolts has increased.

(b) Calculate the critical region for $S^2$

**(3)**

The manager of the factory decides that, in future, he will check each month whether the machine making the bolts is working properly. He uses a 10% level of significance to test whether or not there is evidence that

- the mean length of the bolts has changed
- the variance of the length of the bolts has increased

The next month a random sample of 15 bolts is taken.

The mean length of these bolts is 30.06 mm and the standard deviation is 1.02 mm

(c) With reference to your answers to part (a) and part (b), state whether or not there is any evidence that the machine is **not** working properly.
Give reasons for your answer.

**(2)**

(a) 90% confidence interval: $z = 1.6449$

$30.03 \pm 1.6449 \times \frac{0.868}{\sqrt{15}}$

$= 30.03 \pm 0.368650...$

$= (29.6613..., 30.3986...)$

90% CI for $\mu$: (29.7, 30.4) mm

(b) $H_0: \sigma^2 = 0.868^2$ $\quad H_1: \sigma^2 > 0.868^2$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{14}$ $\quad \chi^2_{14}(0.10) = 21.064$

Reject $H_0$ if $\frac{14S^2}{0.868^2} > 21.064$

$S^2 > \frac{21.064 \times 0.868^2}{14} = 1.13358...$

Critical region: $S^2 > 1.13$

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**

(c) Mean: 30.06 lies inside the confidence interval (29.7, 30.4), so there is no evidence that the mean length of the bolts has changed.

Variance: $s^2 = 1.02^2 = 1.0404$, and $1.0404 < 1.13$ so it is not in the critical region: no evidence that the variance has increased.

So there is no evidence that the machine is not working properly.

**(Total for Question 3 is 8 marks)**

**4.** The random variable $G$ has a continuous uniform distribution over the interval $[-3, 15]$

(a) Calculate $P(G > 12)$

**(1)**

The random variable $H$ has a continuous uniform distribution over the interval $[2, w]$

The random variables $G$ and $H$ are independent and $E(H) = 10$

(b) Show that the probability that $G$ and $H$ are both greater than 12 is $\frac{1}{16}$

**(3)**

The random variable $A$ is the area on a coordinate grid bounded by

$$y = -3$$

$$y = -4|x| + k$$

where $k$ is a value from the continuous uniform distribution over the interval $[5, 10]$

(c) Calculate the expected value of $A$

**(5)**

(a) $P(G > 12) = \frac{15 - 12}{15 - (-3)} = \frac{3}{18} = \frac{1}{6}$

(b) $E(H) = \frac{2 + w}{2} = 10 \Rightarrow w = 18$

$P(H > 12) = \frac{18 - 12}{18 - 2} = \frac{6}{16} = \frac{3}{8}$

G and H independent:

$P(G > 12 \text{ and } H > 12) = \frac{1}{6} \times \frac{3}{8} = \frac{3}{48} = \frac{1}{16}$ as required

(c) Graph of $y = -4|x| + k$: vertex $(0, k)$, gradients $\pm 4$

Triangle above $y = -3$ (see sketch on next page)

$-4|x| + k = -3 \Rightarrow |x| = \frac{k + 3}{4}$

base $= 2 \times \frac{k + 3}{4} = \frac{k + 3}{2}$ height $= k - (-3) = k + 3$

$A = \frac{1}{2} \times \frac{k + 3}{2} \times (k + 3) = \frac{(k + 3)^2}{4}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Question 4 continued**

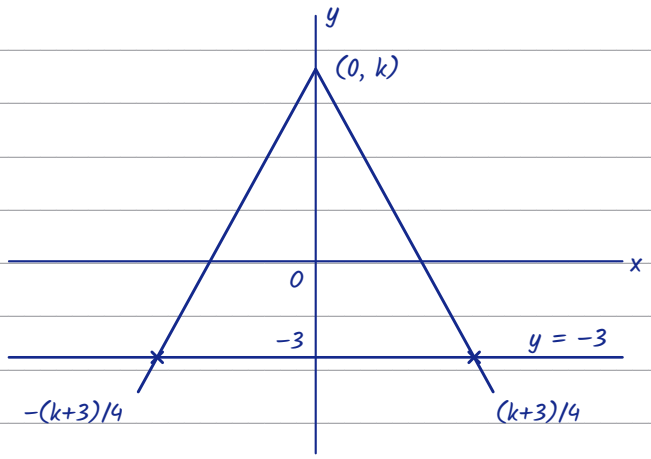
$K \sim U[5, 10]$ so $f(k) = \frac{1}{5}$ for $5 \leqslant k \leqslant 10$

$$E(A) = \int_5^{10} \frac{1}{5} \times \frac{(k+3)^2}{4}\, dk = \frac{1}{20} \int_5^{10} (k+3)^2\, dk$$

$$= \frac{1}{20} \left[ \frac{(k+3)^3}{3} \right]_5^{10} = \frac{1}{60} (13^3 - 8^3)$$

$$= \frac{1}{60} (2197 - 512) = \frac{1685}{60}$$

$$E(A) = \frac{337}{12} \quad (= 28.1)$$




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

**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

**Question 4 continued**

**(Total for Question 4 is 9 marks)**

**5.** A continuous random variable $X$ has probability density function

$$f(x) = \begin{cases} ax^{-2} - bx^{-3} & 2 \leqslant x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where $a$ and $b$ are constants.

Given that $P(X \leqslant 4) = \frac{3}{8}$

(a) use algebraic integration to show that $a = 3$
Show your working clearly.

**(6)**

(b) Find the exact value of the median of $X$

**(4)**

(a) $\int (ax^{-2} - bx^{-3})\,dx = -\frac{a}{x} + \frac{b}{2x^2}$

Total probability: $\int_2^\infty f(x)\,dx = 1$

$\left[ -\frac{a}{x} + \frac{b}{2x^2} \right]_2^\infty = 0 - \left( -\frac{a}{2} + \frac{b}{8} \right) = \frac{a}{2} - \frac{b}{8} = 1$

$\Rightarrow \quad 4a - b = 8 \qquad (1)$

$P(X \leq 4) = \left[ -\frac{a}{x} + \frac{b}{2x^2} \right]_2^4 = \left( -\frac{a}{4} + \frac{b}{32} \right) - \left( -\frac{a}{2} + \frac{b}{8} \right)$

$= \frac{a}{4} - \frac{3b}{32} = \frac{3}{8}$

$\Rightarrow \quad 8a - 3b = 12 \qquad (2)$

(1) × 3: $12a - 3b = 24$

subtract (2): $4a = 12$

$a = 3$ 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

**Question 5 continued**

(b) $b = 4a - 8 = 4 \times 3 - 8 = 4$

$F(m) = \int_2^m (3x^{-2} - 4x^{-3})\,dx = \left[ -\frac{3}{x} + \frac{2}{x^2} \right]_2^m = 0.5$

$-\frac{3}{m} + \frac{2}{m^2} - \left( -\frac{3}{2} + \frac{2}{4} \right) = \frac{1}{2}$

$1 - \frac{3}{m} + \frac{2}{m^2} = \frac{1}{2}$

$\times 2m^2$: $2m^2 - 6m + 4 = m^2$

$m^2 - 6m + 4 = 0$

$m = \frac{6 \pm \sqrt{(36 - 16)}}{2} = \frac{6 \pm \sqrt{20}}{2} = 3 \pm \sqrt{5}$

$3 - \sqrt{5} = 0.76... < 2$, which is outside the range of X, so reject

Median of $X = 3 + \sqrt{5}$

**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

**Question 5 continued**

**(Total for Question 5 is 10 marks)**

**6.** A researcher set up a trial to assess the effect that a food supplement has on the increase in weight of Herdwick lambs. The researcher randomly selected 8 sets of twin lambs. One of each set of twins was given the food supplement and the other had no food supplement. The gain in weight, in kg, of each lamb over the period of the trial was recorded.

| | **Set of twin lambs** | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
|---|---|---|---|---|---|---|---|---|---|
| **Weight gain (kg)** | **With food supplement** | 4.1 | 5.3 | 6.0 | 3.6 | 5.9 | 4.2 | 7.1 | 6.4 |
| | **No food supplement** | 5.0 | 4.8 | 5.2 | 3.4 | 5.1 | 3.9 | 7.0 | 6.5 |

(a) State why a two sample *t*-test is not suitable for use with these data.

**(1)**

(b) Suggest 2 other factors about the lambs that the researcher may need to control when selecting the sample.

**(2)**

(c) State one assumption, in context, that needs to be made for a paired *t*-test to be valid.

**(1)**

For a pair of twin lambs, the random variable $W$ represents the weight gain of the lamb given the food supplement minus the weight gain of the lamb not given the food supplement.

(d) Using the data in the table, calculate a 98% confidence interval for the mean of $W$ Show your working clearly.

**(5)**

The researcher believes that the mean of $W$ is greater than 200 g

(e) Stating your hypotheses clearly, use your confidence interval to explain whether or not there is evidence to support the researcher's belief.

**(3)**

(a) The two samples are not independent: the data are paired, since each pair of results comes from the same set of twin lambs.

(b) The birth weight of the lambs.
The gender of the lambs (and whether or not they were born premature).

(c) The differences between the weight gains of the twin lambs (with supplement – without supplement) are normally distributed.

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 6 continued**

(d) w: −0.9, 0.5, 0.8, 0.2, 0.8, 0.3, 0.1, −0.1

$\Sigma w = 1.7$ $\Sigma w^2 = 2.49$ $\bar{w} = \frac{1.7}{8} = 0.2125$

$s^2 = \frac{2.49 - 8 \times 0.2125^2}{7} = 0.304107...$ $(s = 0.55145...)$

$t_7(1\%) = 2.998$

98% CI: $0.2125 \pm 2.998 \times \sqrt{(0.304107... \div 8)}$

$= 0.2125 \pm 0.584520...$

$= (-0.37202..., 0.79702...)$

98% CI for the mean of W: (−0.372, 0.797) kg

(e) $H_0$: $\mu_W = 0.2$ $H_1$: $\mu_W > 0.2$ (in kg)

200 g = 0.2 kg, and 0.2 lies inside the confidence interval (−0.372, 0.797) (a 98% CI corresponds to a one-tailed test at the 1% level), so do not reject $H_0$.

There is insufficient evidence to support the researcher's belief that the mean of W is greater than 200 g.

**Question 6 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

**Question 6 continued**

**(Total for Question 6 is 12 marks)**

**7.** Two organisations are each asked to carry out a survey to find out the proportion, $p$, of the population that would vote for a particular political party.

The first organisation finds that out of $m$ people, $X$ would vote for this particular political party.

The second organisation finds that out of $n$ people, $Y$ would vote for this particular political party.

An unbiased estimator, $Q$, of $p$ is proposed where

$$Q = k\left(\frac{X}{m} + \frac{Y}{n}\right)$$

(a) Show that $k = \frac{1}{2}$

**(2)**

A second unbiased estimator, $R$, of $p$ is proposed where

$$R = \frac{aX}{m} + \frac{bY}{n}$$

(b) Show that $a + b = 1$

**(2)**

Given that $m = 100$ and $n = 200$ and that $R$ is a better estimator of $p$ than $Q$

(c) calculate the range of possible values of $a$
Show your working clearly.

**(7)**

(a) $X \sim B(m, p)$, $Y \sim B(n, p)$ so $E(X) = mp$, $E(Y) = np$

$E(Q) = k\left(\frac{E(X)}{m} + \frac{E(Y)}{n}\right) = k\left(\frac{mp}{m} + \frac{np}{n}\right) = k(p + p) = 2kp$

Q unbiased: $2kp = p \Rightarrow k = \frac{1}{2}$ as required

(b) $E(R) = \frac{aE(X)}{m} + \frac{bE(Y)}{n} = \frac{amp}{m} + \frac{bnp}{n} = ap + bp$

R unbiased: $ap + bp = p \Rightarrow a + b = 1$ 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

**Question 7 continued**

(c) $\text{Var}(X) = mp(1-p)$, $\text{Var}(Y) = np(1-p)$

$$\text{Var}(Q) = \frac{1}{4}\left(\frac{mp(1-p)}{m^2} + \frac{np(1-p)}{n^2}\right) = \frac{1}{4}p(1-p)\left(\frac{1}{m} + \frac{1}{n}\right)$$

$$\text{Var}(R) = \frac{a^2mp(1-p)}{m^2} + \frac{b^2np(1-p)}{n^2} = p(1-p)\left(\frac{a^2}{m} + \frac{b^2}{n}\right)$$

Both unbiased, so R is better if $\text{Var}(R) < \text{Var}(Q)$:

$$\frac{a^2}{m} + \frac{b^2}{n} < \frac{1}{4}\left(\frac{1}{m} + \frac{1}{n}\right)$$

$m = 100$, $n = 200$, $b = 1 - a$:

$$\frac{a^2}{100} + \frac{(1-a)^2}{200} < \frac{1}{4}\left(\frac{1}{100} + \frac{1}{200}\right) = \frac{3}{800}$$

$\times 800$: $8a^2 + 4(1 - 2a + a^2) < 3$

$12a^2 - 8a + 1 < 0$

$(6a - 1)(2a - 1) < 0$

Critical values: $a = \frac{1}{6}$ or $a = \frac{1}{2}$

$$\frac{1}{6} < a < \frac{1}{2}$$

**Question 7 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

**Question 7 continued**

**(Total for Question 7 is 11 marks)**

**8.** A company packs chickpeas into small bags and large bags.

The weight of a small bag of chickpeas is normally distributed with mean 500 g and standard deviation 5 g

A random sample of 3 small bags of chickpeas is taken.

(a) Find the probability that the total weight of these 3 bags of chickpeas is between 1490 g and 1530 g

**(3)**

The weight of a large bag of chickpeas is normally distributed with mean 1020 g and standard deviation 20 g

One large bag and one small bag of chickpeas are chosen at random.

(b) Calculate the probability that the weight of the large bag of chickpeas is at least 30 g more than twice the weight of the small bag of chickpeas.
Show your working clearly.

**(6)**

(a) $T = S_1 + S_2 + S_3$

$E(T) = 3 \times 500 = 1500 \qquad Var(T) = 3 \times 5^2 = 75$

$T \sim N(1500, 75)$

$$P(1490 < T < 1530) = P\left(\frac{1490 - 1500}{\sqrt{75}} < Z < \frac{1530 - 1500}{\sqrt{75}}\right)$$

$= P(-1.1547... < Z < 3.4641...)$

$= 0.99973... - 0.12410... = 0.87562...$

$= 0.876$

(b) Require $P(L \geqslant 2S + 30)$. Let $W = L - 2S - 30$

$E(W) = 1020 - 2 \times 500 - 30 = -10$

$Var(W) = Var(L) + 2^2 Var(S) = 20^2 + 4 \times 5^2 = 500$

$W \sim N(-10, 500)$

$$P(W \geqslant 0) = P\left(Z \geqslant \frac{0 - (-10)}{\sqrt{500}}\right) = P(Z \geqslant 0.44721...)$$

$= 1 - 0.67263... = 0.32736...$

$= 0.327$

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 8 continued**

**Question 8 continued**

**(Total for Question 8 is 9 marks)**

**TOTAL FOR PAPER IS 75 MARKS**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA