1. A woodwork teacher measures the width, w mm, of a board. The measured width, x mm, is normally distributed with mean w mm and standard deviation 0.5 mm.

(a) Find the probability that x is within 0.6 mm of w.

(2)

The same board is measured 16 times and the results are recorded.

(b) Find the probability that the mean of these results is within 0.3 mm of w.

(4)

Given that the mean of these 16 measurements is 35.6 mm,

(c) find a 98% confidence interval for w.

(4)
(Total 10 marks)

2. The heights of a random sample of 10 imported orchids are measured. The mean height of the sample is found to be 20.1 cm. The heights of the orchids are normally distributed.

Given that the population standard deviation is 0.5 cm,

(a) estimate limits between which 95% of the heights of the orchids lie,

(3)

(b) find a 98% confidence interval for the mean height of the orchids.

(4)

A grower claims that the mean height of this type of orchid is 19.5 cm.

(c) Comment on the grower's claim. Give a reason for your answer.

(2)
(Total 9 marks)

3. A sample of size 8 is to be taken from a population that is normally distributed with mean 55 and standard deviation 3. Find the probability that the sample mean will be greater than 57.

(Total 5 marks)

4. The weights of adult men are normally distributed with a mean of 84 kg and a standard deviation of 11 kg.

(a) Find the probability that the total weight of 4 randomly chosen adult men is less than 350 kg.

(5)

The weights of adult women are normally distributed with a mean of 62 kg and a standard deviation of 10 kg.

(b) Find the probability that the weight of a randomly chosen adult man is less than one and a half times the weight of a randomly chosen adult woman.

(6)
(Total 11 marks)

5. A random sample $X_1, X_2, \ldots, X_{10}$ is taken from a normal population with mean 100 and standard deviation 14.

(a) Write down the distribution of $\overline{X}$, the mean of this sample.

(2)

(b) Find $\mathrm{P}(|\overline{X} - 100| > 5)$.

(3)
(Total 5 marks)

6. A random sample of the invoices, for books purchased by the customers of a large bookshop, was classified by book cover (hardback, paperback) and type of book (novel, textbook, general interest). As part of the analysis of these invoices, an approximate χ^2 statistic was calculated and found to be 11.09.

Assuming that there was no need to amalgamate any of the classifications, carry out an appropriate test to determine whether or not there was any association between book cover and type of book. State your hypotheses clearly and use a 5% level of significance.

(Total 6 marks)

7. As part of a research project into the role played by cholesterol in the development of heart disease a random sample of 100 patients was put on a special fish-based diet. A different random sample of 80 patients was kept on a standard high-protein low-fat diet. After several weeks their blood cholesterol was measured and the results summarised in the table below.

Group	Sample size	Mean drop in cholesterol (mg/dl)	Standard deviation
Special diet	100	75	22
Standard diet	80	64	31

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the special diet is more effective in reducing blood cholesterol levels than the standard diet.

(9)

(b) Explain briefly any assumptions you made in order to carry out this test.

(2)

(Total 11 marks)

8. Breakdowns on a certain stretch of motorway were recorded each day for 80 consecutive days. The results are summarised in the table below.

Number of breakdowns	0	1	2	> 2
Frequency	38	32	10	0

It is suggested that the number of breakdowns per day can be modelled by a Poisson distribution.

Using a 5% level of significance, test whether or not the Poisson distribution is a suitable model for these data. State your hypotheses clearly.

(Total 13 marks)

9. The random variable R is defined as $R = X + 4Y$ where $X \sim N(8, 2^2)$, $Y \sim N(14, 3^2)$ and X and Y are independent.

Find

(a) E(R),

(2)

(b) Var (R),

(3)

(c) P(R < 41)

(3)

The random variables Y_1, Y_2 and Y_3 are independent and each has the same distribution as Y. The random variable S is defined as

$$S = \sum_{i=1}^{3} Y_i - \tfrac{1}{2}X .$$

(d) Find Var (S).

(4)

(Total 12 marks)

10. As part of her statistics project, Deepa decided to estimate the amount of time A-level students at her school spend on private study each week. She took a random sample of students from those studying Arts subjects, Science subjects and a mixture of Arts and Science subjects. Each student kept a record of the time they spent on private study during the third week of term.

(a) Write down the name of the sampling method used by Deepa.

(1)

(b) Give a reason for using this method and give one advantage this method has over simple random sampling.

(2)

The results Deepa obtained are summarised in the table below.

Type of student	Sample size	Mean number of hours
Arts	12	12.6
Science	12	14.1
Mixture	8	10.2

(c) Show that an estimate of the mean time spent on private study by A level students at Deepa's school, based on these 32 students is 12.56, to 2 decimal places.

(3)

The standard deviation of the time spent on private study by students at the school was 2.48 hours.

(d) Assuming that the number of hours spent on private study is normally distributed, find a 95% confidence interval for the mean time spent on private study by A level students at Deepa's school.

(4)

A member of staff at the school suggested that A level students should spend on average 12 hours each week on private study.

(e) Comment on this suggestion in the light of your interval.

(2)
(Total 12 marks)

11. For one of the activities at a gymnastics competition, 8 gymnasts were awarded marks out of 10 for each of artistic performance and technical ability. The results were as follows.

Gymnast	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Technical ability	8.5	8.6	9.5	7.5	6.8	9.1	9.4	9.2
Artistic performance	6.2	7.5	8.2	6.7	6.0	7.2	8.0	9.1

The value of the product moment correlation coefficient for these data is 0.774.

(a) Stating your hypotheses clearly and using a 1% level of significance, interpret this value.
(5)

(b) Calculate the value of the rank correlation coefficient for these data.
(6)

(c) Stating your hypotheses clearly and using a 1% level of significance, interpret this coefficient.
(3)

(d) Explain why the rank correlation coefficient might be the better one to use with these data.
(2)
(Total 16 marks)

1. (a) $E \sim \text{N}(0, 0.5^2)$

or

$X \sim \text{N}\left(w, 0.5^2\right)$

$\text{P}(|E| < 0.6) \quad = \text{P}\left(|Z| < \frac{0.6}{0.5}\right)$

or

$\text{P}\left(|X - w| < 0.6\right) = \text{P}\left(|Z| < \frac{0.6}{0.5}\right)$ M1

$= \text{P}(|Z| < 1.2)$

$= 2 \times 0.8849 - 1 = 0.7698$ awrt **0.770** A1 2

Note

1st M1 for identifying a correct probability (they must have the 0.6) and attempting to standardise. Need | |. This mark can given for 0.8849 – 0.1151 seen as final answer.

1st A1 for awrt 0.770. NB an answer of 0.3849 or 0.8849 scores M0A0 (since it implies no | |)

M1 may be implied by a correct answer

(b) $\bar{E} \sim \text{N}\left(0, \frac{1}{64}\right)$

or

$\bar{X} \sim \text{N}\left(w, \frac{0.5^2}{16}\right)$ M1

$\text{P}\left(|\bar{E}| < 0.3\right) \quad = \text{P}\left(|Z| < \frac{0.3}{\frac{1}{8}}\right)$

or

$\text{P}\left(|\bar{X} - w| < 0.3\right) = \text{P}\left(|Z| < \frac{0.3}{\frac{1}{8}}\right)$ M1, A1

$= \text{P}(|Z| < 2.4)$

$= 2 \times 0.9918 - 1 = 0.9836$ awrt **0.984** A1 4

Note

1st M1 for a correct attempt to define $\bar{E}$ or $\bar{X}$ but must attempt $\frac{\sigma^2}{n}$. Condone labelling as E or X

This mark may be implied by standardisation in the next line.

2nd M1 for identifying a correct probability statement using $\bar{E}$ or $\bar{X}$. Must have 0.3 and | |

1st A1 for correct standardisation as printed or better

2nd A1 for awrt 0.984

The M marks may be implied by a correct answer.

Sum of 16, not means

1st M1 for correct attempt at suitable sum distribution with correct variance (= 16 × $\frac{1}{4}$)

2nd M1 for identifying a correct probability. Must have 4.8 and | |

1st A1 for correct standardisation i.e. need to see $\frac{4.8}{\sqrt{4}}$ or better

(c) $35.6 \pm 2.3263 \times \frac{1}{8}$ M1 B1

(**35.3, 35.9**) A1, A1 4

<u>**Note**</u>

M1 for $35.6 \pm z \times \frac{0.5}{\sqrt{16}}$

B1 for 2.3263 or better. Use of 2.33 will lose this mark but can still score ¾

1st A1 for awrt 35.3

2nd A1 for awrt 35.9

[10]

2. (a) Limits are 20.1 ± 1.96 × 0.5 M1 B1

<u>**(19.1, 21.1)**</u> A1cso 3

<u>**Note**</u>

M1 for 20.1 ± z × 0.5. Need 20.1 and 0.5 in correct places with no $\sqrt{10}$

B1 for z = 1.96 (or better)

A1 for awrt 19.1 <u>and</u> awrt 21.1
but must have scored both M1 and B1

[Correct answer only scores 3/3]

(b) 98% confidence limits are M1

$20.1 \pm 2.3263 \times \frac{0.5}{\sqrt{10}}$ B1

<u>**(19.7, 20.5)**</u> A1A1 4

<u>**Note**</u>

M1 for $20.1 \pm z \times \frac{0.5}{\sqrt{10}}$ need to see 20.1, 0.5

and $\sqrt{10}$ in correct places

B1 for z = 2.3263 (or better)

1st A1 for awrt 19.7

2nd A1 for awrt 20.5

[Correct answer only scores M1B0A1A1]

(c)	The growers claim is not correct	B1	
	Since 19.5 does not lie in the interval (19.7, 20.5)	dB1	2

<u>Note</u>

1st B1 for rejection of the claim. Accept "unlikely" or "not correct"

2nd dB1 Dependent on scoring 1st B1 in this part for rejecting grower's claim for an argument that supports this. Allow comment on <u>their</u> 98% CI from (b)

[9]

3.	$X \sim \text{N}(55, 3^2)$ therefore $\bar{X} \sim \text{N}(55, \frac{9}{8})$		B1 B1	
	$\text{P}(\bar{X} > 57) = \text{P}\left(Z > \frac{57-55}{\sqrt{\frac{9}{8}}}\right) = \text{P}(Z > 1.8856\ldots)$		M1	
	$= 1 - 0.9706$		M1	
	$= 0.0294$	**<u>0.0294~0.0297</u>**	A1	5

–

<u>Note</u>

1st B1 for $\bar{X}$ ~ normal and μ = 55, may be implied but must be $\bar{X}$

2nd B1 for Var($\bar{X}$) or st. dev of $\bar{X}$ e.g. $\bar{X} \sim \text{N}(55, \frac{9}{8})$ or $\bar{X} \sim \text{N}\left(55, \left(\frac{3}{\sqrt{8}}\right)^2\right)$ for B1B1

Condone use of X if they clearly mean $\bar{X}$ so $X \sim \text{N}\left(55, \frac{9}{8}\right)$ is OK for B1B1

1st M1 for an attempt to standardize with 57 and mean of 55 and their st. dev. $\neq 3$

2nd M1 for 1 – tables value. Must be trying to find a probability < 0.5

A1 for answers in the range 0.0294~0.0297

<u>ALT</u>

$$\sum_{1}^{8} X_i \sim \text{N}\left(8\times 55, 8\times 3^2\right)$$

1st B1 for $\sum X \sim$ normal and mean = 8×55

2nd B1 for variance = 8×3^2

1st M1 for attempt to standardise with 57 × 8, mean of 55 × 8 and their st dev ≠ 3

[5]

4.	(a)	$X = M_1 + M_2 + M_3 + M_4 \sim \text{N}(336, 22^2)$	μ = **336**	B1	
			$\sigma^2 = 22^2$ or **484**	B1	
		$\text{P}(X < 350) = \text{P}\left(Z < \dfrac{350-336}{22}\right)$		M1	
		$= \text{P}(Z < 0.64)$	awrt **0.64**	A1	
		=	awrt **0.738** or **0.739**	A1	5

2nd B1 for $\sigma = 22$ or $\sigma^2 = 22^2$ or 484

M1 for standardising with their mean and standard deviation (ignore direction of inequality)

	(b)	$M \sim \text{N}(84, 121)$ and $W \sim N(62, 100)$ Let $Y = M - 1.5W$		M1	
		$\text{E}(Y) = 84 - 1.5 \times 62 = -9$		A1	
		$\text{Var}(Y) = \text{Var}(M) + 1.5^2\ \text{Var}(W)$		M1	
		$= 11^2 + 1.5^2 \times 10^2 = 346$		A1	
		$\text{P}(Y < 0), = \text{P}(Z < 0.48\ldots) =$	awrt **0.684 ~ 0.686**	B1	6

1st M1 for attempting to find Y. Need to see $\pm(M - 1.5W)$ or equiv. May be implied by Var(Y).

1st A1 for a correct value for their E(Y) i.e. usually ± 9. Do not give M1A1 for a “lucky” ± 9.

2nd M1 for attempting Var(Y) e.g. ... + $1.5^2 \times 10^2$ or $11^2 + 1.5^2 \times$ ...

3rd M1 for attempt to calculate the correct probability. Must be attempting a probability > 0.5. Must attempt to standardise with a relevant mean and standard deviation

Using $\sigma^2_M = 11$ or $\sigma^2_W = 10$ is not a misread.

[11]

5.	(a)	$\bar{X} \sim \text{N}\left(100, \frac{14^2}{10}\right)$		
		Normal	B1	
		$100,\ \frac{14^2}{10}$	B12	
	(b)	$\text{P}(\lvert \bar{X} - 100 \rvert > 5) = \text{P}(\bar{X} > 105) + \text{P}(\bar{X} < 95)$	M1	
		$= 2\text{P}(\bar{X} > 105)$		
		$= 2\text{P}\left(Z > \frac{105-100}{\sqrt{\frac{14^2}{10}}}\right)$	A1	
		$= 2\text{P}(Z > 1.13)$		
		$= 0.2584$	A1	3

[5]

6.	H_0: No association between type and cover		
	H_1: Association between type and cover	both B1	
	$\alpha = 0.05;\ \nu = 2;$	M1 A1	
	Critical value = 5.991	B1	
	$\sum \frac{(O-E)^2}{E} = 11.09$		
	Since 11.09 is in the critical region, there is evidence of association between type of book and type of cover	M1 A1	6

[6]

7.	(a)	H_0: $\mu_{sp} = \mu_{st}$; H_1: $\mu_{sp} > \mu_{st}$;	B1 B1	
		$\alpha = 0.05$; critical region: $z > 1.6449$	B1	
		standard error = $\sqrt{\frac{22^2}{100}+\frac{31^2}{80}} = 4.1051\ldots$		
		$z = \frac{75-64}{4.1051...} = 2.68$	M1 A1	
		Since 2.68 is in the critical region there is evidence to reject H_0 and conclude that the special diet is more effective in reducing blood cholesterol.	M1 A1 ft	9
	(b)	Drop in blood cholesterol levels are normally distributed, or Central Limit Theorem can be applied, or standard deviations of the populations are 22 and 31		
		Any two	B1 B12	

[11]

8.	(a)	H_0: Poisson distribution is a suitable model H_1: Poisson distribution is not a suitable model	both B1	
		From these data $\lambda = \frac{52}{80} = 0.65$	M1 A1	
		Expected frequencies 41.76, 27.15, $\underbrace{8.82, 2.27}_{11.09}$		
		80 × P(X = x)	M1 A2/1/0	
		Amalgamation	M1	
		$\alpha = 0.05$, $\nu = 3 - 1 - 1 = 1$; critical value = 3.841	B1 ft; B1 ft	
		$\sum\frac{(O-E)^2}{E} = 1.312$	M1 A1 ft	
		Since 1.312 is not the critical region there is insufficient evidence to reject H_0 and we can conclude that the Poisson model is a suitable one.	M1 A1 ft	13

[13]

9.	(a)	$E(R) = E(X) + 4E(Y) = 8 + (4 \times 14) = 64$	M1 A1	2
	(b)	$\text{Var}(R) = \text{Var}(X) + 16\,\text{Var}(Y) = 2^2 + (16 \times 3^2)$	M1 A1	
		$= 148$	A1	3

	(c)	$P(R < 41) = P\left(Z < \frac{41 - 64}{\sqrt{148}}\right) = P(Z < -1.89)$	M1 A1 ft	
		$= 0.0294$	A1	3
	(d)	$\text{Var}(S) = 3\,\text{Var}(Y) + (\tfrac{1}{2})^2\,\text{Var}(X)$	M1 M1	
		$= 27 + 1$	A1	
		$= 28$	A1	4

[12]

10.	(a)	Stratified sampling	B1	1
	(b)	Uses naturally occurring (strata) groupings e.g. variance of estimator of population mean is usually reduced, individual strata estimates available	either B1 B1	2
	(c)	$\bar{x} = \frac{(12 \times 12.6) + (12 \times 14.1) + (8 \times 10.2)}{32}$	M1 A1	
		$= 12.56$	A1	3
	(d)	Confidence interval is $12.56 \pm 1.96 \times \frac{2.48}{\sqrt{32}}$	1.96 M1	
		i.e. $12.56 \pm 0.859276\ldots$	B1	
		i.e. (11.70, 13.42) *accept (11.7, 13.4)*	A14	
	(e)	12 is within the confidence interval; so the time spent by these students is in agreement with the suggestion of the member of staff.	B1; B1	2

[12]

11.	(a)	$H_0: \rho = 0$, $H_1: \rho > 0$	B1 B1	
		$\alpha = 0.01$, critical value = 0.7887	B1	
		Since 0.774 is not in the critical region there is insufficient evidence of positive correlation.	M1 A1	5

(b) e.g.

R_T	3	4	8	2	1	5	7	6
R_A	2	5	7	3	1	4	6	8

Ranks	M1	
All correct	A1	
$\sum d^2 = 10$	M1 A1	
$r_s = 1 - \frac{6 \times 10}{8 \times 63} = 0.881$	M1 A1	6
(c) $H_0: \rho = 0$, $H_1: \rho > 0$	both B1	
$\alpha = 0.01$; critical value: 0.8333	B1	
Since 0.881 is in the critical region there is evidence of positive correlation.	A1	3
(d) Because it makes no distributional assumptions about the data or order is more important than the mark	B1	
Product moment correlation assumes bivariate normality and it is very unlikely that these scores will be distributed this way.	B1	2

[16]

1. The first two parts of this question proved quite challenging for many candidates who were unable to handle the modulus. Many simply missed the significance of the phrase "…is within…" and others misinterpreted the ranges dividing by 2. Those who did interpret the question correctly usually had few problems in calculating the required probabilities although there were some errors with the variance in part (b).

The confidence interval in part (c) was answered very well by most candidates and many correct solutions were seen. Few failed to use z = 2.3263 and only a small minority used an incorrect standard error.

2. It is often the case when we set questions like part (a) that many candidates simply calculate a 95% confidence interval for the mean and the same happened here. They then repeated the technique in part (b), usually with the correct z value, and there were many fully correct confidence intervals found here. Some candidates were not prepared to make a decision in part (c) and hedged their comments, possibly because 19.5 was "in" the answer to part (a) but not in the confidence interval in (b). The examiners required a clear rejection of the grower's claim, based on the fact that 19.5 was outside the 98% confidence interval and a good number of candidates gained both marks for these two simple statements.

3. This was a very straightforward question and most gained full marks. Sometimes their notation was far from clear. It would be nice to see all candidates writing $X \sim \mathrm{N}(55,3^2)$ and then stating $\bar{X} \sim \mathrm{N}\left(55, \frac{3^2}{8}\right)$ before calculating $\mathrm{P}(\bar{X} > 57)$ but it was clear from their calculations that they knew what they were doing even if they did not always communicate their intentions clearly.

4. In part (a) most found the mean of 336 but some thought the variance was 44^2 however many did obtain the correct values here and went on to complete this part successfully. Part (b) though proved more challenging. Some never defined a suitable random variable such as $M - 1.5W$ and made no progress. Those who did make a suitable start were usually able to obtain the mean of –9 but the variance calculation caused difficulties with many using 1.5 instead of 1.5^2. The standardisation was usually carried out correctly but a diagram would have helped some get their final probability the right way round.

5. No Report available for this question.

6. No Report available for this question.

7. No Report available for this question.

8. No Report available for this question.

9. No Report available for this question.

10. No Report available for this question.

11. No Report available for this question.