1. The number of goals scored by a football team is recorded for 100 games. The results are summarised in Table 1 below.

Number of goals	Frequency
0	40
1	33
2	14
3	8
4	5

Table 1

(a) Calculate the mean number of goals scored per game.

(2)

The manager claimed that the number of goals scored per match follows a Poisson distribution. He used the answer in part (a) to calculate the expected frequencies given in Table 2.

Number of goals	Expected Frequency
0	34.994
1	r
2	s
3	6.752
≥ 4	2.221

Table 2

(b) Find the value of r and the value of s giving your answers to 3 decimal places.

(3)

(c) Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim.

(7)
(Total 12 marks)

2. 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 P($|\overline{X} - 100| > 5$).

(3)
(Total 5 marks)

3. 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)

4. 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)

5. 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)

6. 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)

7. 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)

8. 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) $\lambda = \dfrac{0\times 40 + 1\times 33 + 2\times 14 + 3\times 8 + 4\times 5}{100} = 1.05$ M1 A1 2

Note

M1 for an attempt to find the mean– at least 2 terms on numerator seen Correct answer only will score both marks

(b) Using Expected frequency = $100 \times P(X = x)$

$= 100 \times \dfrac{e^{-1.05}1.05^x}{x!}$ gives M1

$r = 36.743$ awrt 36.743 or 36.744 A1

$s = 19.290$ 19.29 or awrt 19.290 A1 3

Note

M1 for use of correct formula (ft their mean).
1st A1 for r, 2nd A1 for s (19.29 OK)

(c) H_0 : Poisson distribution is a suitable model B1

H_1 : Poisson distribution is not a suitable model

Number of goals	Frequency	Expected frequency	
0	40	34.994	
1	33	36.743	
2	14	19.290	
3	8	6.752	8.972443
≥ 4	5	2.221	

M1

$\upsilon = 4 - 1 - 1 = 2$ B1ft

CR : χ^2_2 (0.05) > 5.991 B1

$$\sum \frac{(O-E)^2}{E} = \frac{(40-34.9937)^2}{34.9937} + ...$$
$$+ \frac{(13-8.972443)^2}{8.972443}$$ M1

[=0.7161…+0.3813…
+1.4508…+1.80789..] A1

= 4.356. (ans in range 4.2 – 4.4)

Not in critical region

Number of goals scored can follow a Poisson distribution / managers claim is justified A1 ft 7

Note

1st B1 Must have both hypotheses and mention Poisson at least once inclusion of their value for mean in hypotheses is B0 but condone in conclusion

1st M1 for an attempt to pool ≥ 4

2nd B1ft for $n - 1 - 1 = 2$ i.e realising that they must subtract 2 from their n

3rd B1 for 5.991 only

2nd M1 for an attempt at the test statistic, at least 2 correct expressions/values (to 3sf)

1st A1 for answers in the range 4.2~4.4

2nd A1 for correct comment in context based on their test statistic and their cv that mentions goals or manager. Dependent on 2nd M1

Condone mention of Po(1.05) in conclusion

Score A0 for inconsistencies e.g. "significant" followed by "manager's claim is justified"

[12]

2. (a) $\overline{X} \sim \mathrm{N}\left(100, \frac{14^2}{10}\right)$

Normal B1

100, $\frac{14^2}{10}$ B12

(b) $\mathrm{P}(|\overline{X} - 100| > 5) = \mathrm{P}(\overline{X} > 105) + \mathrm{P}(\overline{X} < 95)$ M1

$= 2\mathrm{P}(\overline{X} > 105)$

$= 2\mathrm{P}\left(Z > \frac{105 - 100}{\sqrt{\frac{14^2}{10}}}\right)$ A1

$= 2\mathrm{P}(Z > 1.13)$

$= 0.2584$ A1 3

[5]

3. 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]

4. (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]

5. (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]

6.	(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)	$\text{Var}(S) = 3\ \text{Var}(Y) + (\frac{1}{2})^2\ \text{Var}(X)$	M1 M1	
		$= 27 + 1$	A1	
		$= 28$	A1	4

[12]

7.	(a)	Stratified sampling	B1	1
	(b)	Uses naturally occurring (strata) groupings		
		e.g. variance of estimator of population mean is usually	either B1	
		reduced, individual strata estimates available	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]

8. (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. Parts (a) and (b) were answered very well and most scored full marks on these two parts but part (c) proved more challenging. Many insisted on including the mean of 1.05 in their hypotheses even though this was incompatible with their correct treatment of the degrees of freedom. The pooling of the last two groups was usually carried out and the calculation of the test statistic was often correct. There was some confusion over the calculation of the degrees of freedom though: many subtracted 2 but others only 1 and some were not sure whether to subtract from the number of classes before or after the pooling. A number failed to score the final mark because their conclusion was not given in context: comments such as "there is evidence to support the manager's claim" or "there is evidence that the number of goals scored in football matches does follow a Poisson distribution" are fine; "the data follows a Poisson distribution" is not.

2. No Report available for this question.

3. No Report available for this question.

4. No Report available for this question.

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.