Paper Reference(s)

6683

Edexcel GCE

Statistics S1

Advanced Subsidiary

Specimen Paper

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Answer Book (AB16)	Nil
Mathematical Formulae (Lilac)	
Graph Paper (ASG2)	

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.
When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has eight questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. (*a*) Explain what you understand by a statistical model.

(2)

(*b*) Write down a random variable which could be modelled by

(i) a discrete uniform distribution,

(ii) a normal distribution.

(2)

---

2. A group of students believes that the time taken to travel to college, T minutes, can be assumed to be normally distributed. Within the college 5% of students take at least 55 minutes to travel to college and 0.1% take less than 10 minutes.

Find the mean and standard deviation of T.

(9)

---

3. The discrete random variable X has probability function

$$P(X = x) = \begin{cases} kx, & x = 1, 2, 3, 4, 5, \\ 0, & \text{otherwise.} \end{cases}$$

(*a*) Show that $k = \frac{1}{15}$.

(3)

Find the value of

(*b*) $E(2X + 3)$,

(5)

(*c*) $Var(2X - 4)$.

(6)

---

4. A drilling machine can run at various speeds, but in general the higher the speed the sooner the drill needs to be replaced. Over several months, 15 pairs of observations relating to speed, s revolutions per minute, and life of drill, h hours, are collected.

For convenience the data are coded so that $x = s - 20$ and $y = h - 100$ and the following summations obtained.

$\Sigma x = 143$; $\Sigma y = 391$; $\Sigma x^2 = 2413$; $\Sigma y^2 = 22441$; $\Sigma xy = 484$.

(*a*) Find the equation of the regression line of h on s.

(10)

(*b*) Interpret the slope of your regression line.

(2)

Estimate the life of a drill revolving at 30 revolutions per minute.

(2)

5. (*a*) Explain briefly the advantages and disadvantages of using the quartiles to summarise a set of data.

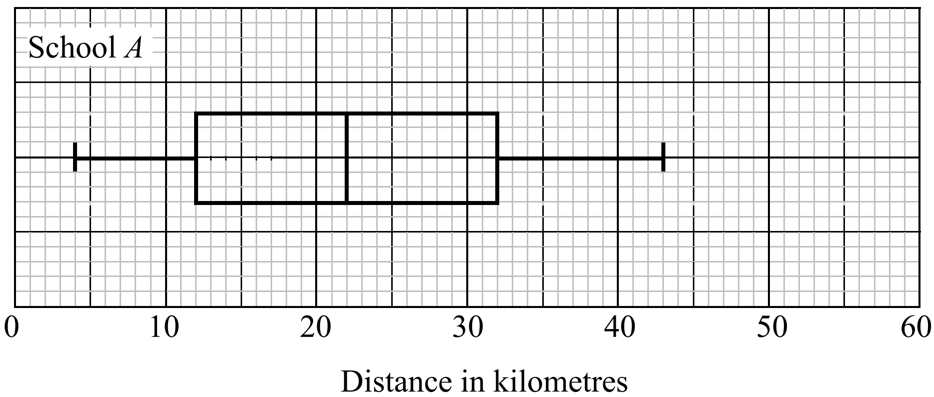
(4)

(*b*) Describe the main features and uses of a box plot.

(3)

The distances, in kilometres, travelled to school by the teachers in two schools, *A* and *B*, in the same town were recorded. The data for School *A* are summarised in Diagram 1.

Diagram 1



For School *B*, the least distance travelled was 3 km and the longest distance travelled was 55 km. The three quartiles were 17, 24 and 31 respectively.

An outlier is an observation that falls either 1.5 × (interquartile range) above the upper quartile or 1.5 × (interquartile range) below the lower quartile.

(*c*) Draw a box plot for School *B*.

(5)

(*d*) Compare and contrast the two box plots.

(4)

6. For any married couple who are members of a tennis club, the probability that the husband has a degree is $\frac{3}{5}$ and the probability that the wife has a degree is $\frac{1}{2}$. The probability that the husband has a degree, given that the wife has a degree, is $\frac{11}{12}$.

A married couple is chosen at random.

(*a*) Show that the probability that both of them have degrees is $\frac{11}{24}$.

(2)

(*b*) Draw a Venn diagram to represent these data.

(5)

Find the probability that

(*c*) only one of them has a degree,

(2)

(d) neither of them has a degree.

(3)

Two married couples are chosen at random.

(*e*) Find the probability that only one of the two husbands and only one of the two wives have degrees.

(6)

END

Paper Reference(s)

6684

Edexcel GCE

Statistics S2

Advanced Level

Specimen Paper

Time: 1 hour 30 minutes

Materials required for examination papers	Items included with question
Answer Book (AB16)	Nil
Mathematical Formulae (Lilac)	
Graph Paper (ASG2)	

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S2), the paper reference (6684), your surname, other name and signature.
When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has seven questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. A school held a disco for years 9, 10 and 11 which was attended by 500 pupils. The pupils were registered as they entered the disco. The disco organisers were keen to assess the success of the event. They designed a questionnaire to obtain information from those who attended.

(*a*) State one advantage and one disadvantage of using a sample survey rather than a census.

(2)

(*b*) Suggest a suitable sampling frame.

(1)

(*c*) Identify the sampling units.

(1)

---

2. A piece of string AB has length 12 cm. A child cuts the string at a randomly chosen point P, into two pieces. The random variable X represents the length, in cm, of the piece AP.

(*a*) Suggest a suitable model for the distribution of X and specify it fully

(2)

(*b*) Find the cumulative distribution function of X.

(4)

(*c*) Write down $P(X < 4)$.

(1)

---

3. A manufacturer of chocolates produces 3 times as many soft centred chocolates as hard centred ones.

Assuming that chocolates are randomly distributed within boxes of chocolates, find the probability that in a box containing 20 chocolates there are

(*a*) equal numbers of soft centred and hard centred chocolates,

(3)

(*b*) fewer than 5 hard centred chocolates.

(2)

A large box of chocolates contains 100 chocolates.

(*c*) Write down the expected number of hard centred chocolates in a large box.

(2)

---

4. A company director monitored the number of errors on each page of typing done by her new secretary and obtained the following results:

No. of errors	0	1	2	3	4	5
No. of pages	37	65	60	49	27	12

(*a*) Show that the mean number of errors per page in this sample of pages is 2.
(2)

(*b*) Find the variance of the number of errors per page in this sample.
(2)

(*c*) Explain how your answers to parts (*a*) and (*b*) might support the director's belief that the number of errors per page could be modelled by a Poisson distribution.
(1)

Some time later the director notices that a 4-page report which the secretary has just typed contains only 3 errors. The director wishes to test whether or not this represents evidence that the number of errors per page made by the secretary is now less than 2.

(*d*) Assuming a Poisson distribution and stating your hypothesis clearly, carry out this test. Use a 5% level of significance.
(6)

---

5. In Manuel's restaurant the probability of a customer asking for a vegetarian meal is 0.30. During one particular day in a random sample of 20 customers at the restaurant 3 ordered a vegetarian meal.

(a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the proportion of vegetarian meals ordered that day is unusually low.
(5)

Manuel's chef believes that the probability of a customer ordering a vegetarian meal is 0.10. The chef proposes to take a random sample of 100 customers to test whether or not there is evidence that the proportion of vegetarian meals ordered is different from 0.10.

(*b*) Stating your hypotheses clearly, use a suitable approximation to find the critical region for this test. The probability for each tail of the region should be as close as possible to 2.5%.
(6)

(*c*) State the significance level of this test giving your answer to 2 significant figures.
(1)

---

6. A biologist is studying the behaviour of sheep in a large field. The field is divided up into a number of equally sized squares and the average number of sheep per square is 2.25. The sheep are randomly spread throughout the field.

(*a*) Suggest a suitable model for the number of sheep in a square and give a value for any parameter or parameters required.

(1)

Calculate the probability that a randomly selected sample square contains

(*b*) no sheep,

(1)

(*c*) more than 2 sheep.

(4)

A sheepdog has been sent into the field to round up the sheep.

(*d*) Explain why the model may no longer be applicable.

(1)

In another field, the average number of sheep per square is 20 and the sheep are randomly scattered throughout the field.

(*e*) Using a suitable approximation, find the probability that a randomly selected square contains fewer than 15 sheep.

(7)

---

7. The continuous random variable X has probability density function f(x) given by

$$\mathrm{f}(x)=\begin{cases}\frac{1}{20}x^3, & 1\le x\le 3\\ 0, & \text{otherwise}\end{cases}$$

(a) Sketch f(x) for all values of x.

(3)

(b) Calculate E(X).

(3)

(c) Show that the standard deviation of X is 0.459 to 3 decimal places.

(3)

(d) Show that for $1 \le x \le 3$, P($X \le x$) is given by $\frac{1}{80}\left(x^4-1\right)$ and specify fully the cumulative distribution function of X.

(5)

(e) Find the interquartile range for the random variable X.

(4)

Some statisticians use the following formula to estimate the interquartile range:

$$\text{interquartile range} = \frac{4}{3}\times\text{standard deviation.}$$

(f) Use this formula to estimate the interquartile range in this case, and comment.

(2)

END

Paper Reference(s)

6670

Edexcel GCE

Statistics S3

Advanced Level

Specimen Paper

Time: 1 hour 30 minutes

Materials required for examination papers	**Items included with question**
Answer Book (AB16) Mathematical Formulae (Lilac) Graph Paper (ASG2)	Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6670), your surname, other name and signature.
When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has six questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. The 240 members of a bowling club are listed alphabetically in the club's membership book. The committee wishes to select a sample of 30 members to fill in a questionnaire about the facilities the club offers.

(a) Explain how the committee could use a table of random numbers to take a systematic sample.

(3)

(b) Give one advantage of this method over taking a simple random sample.

(1)

---

2. The weights of pears, P grams, are normally distributed with a mean of 110 and a standard deviation of 8. Geoff buys a bag of 16 pears.

(a) Write down the distribution of $\overline{P}$, the mean weight of the 16 pears.

(2)

(b) Find $\mathrm{P}(110 < \overline{P} < 113)$.

(3)

---

3. The three tasks most frequently carried out in a garage are A, B and C. For each of the tasks the times, in minutes, taken by the garage mechanics are assumed to be normally distributed with means and standard deviations given in the following table.

Task	Mean	Standard deviation
A	225	38
B	165	23
C	185	27

Assuming that the times for the three tasks are independent, calculate the probability that

(a) the total time taken by a single randomly chosen mechanic to carry out all three tasks lies between 533 and 655 minutes,

(5)

(b) a randomly chosen mechanic takes longer to carry out task B than task C.

(5)

---

4. At the end of a season a league of eight ice hockey clubs produced the following table showing the position of each club in the league and the average attendances (in hundreds) at home matches.

Club	A	B	C	D	E	F	G	H
Position	1	2	3	4	5	6	7	8
Average	37	38	19	27	34	26	22	32

(a) Calculate the Spearman rank correlation coefficient between position in the league and average home attendance.

(5)

(b) Stating clearly your hypotheses and using a 5% two-tailed test, interpret your rank correlation coefficient.

(4)

Many sets of data include tied ranks.

(c) Explain briefly how tied ranks can be dealt with.

(2)

---

5. For a six-sided die it is assumed that each of the sides has an equal chance of landing uppermost when the die is rolled.

(a) Write down the probability function for the random variable X, the number showing on the uppermost side after the die has been rolled.

(2)

(b) State the name of the distribution.

(1)

A student wishing to check the above assumption rolled the die 300 times and for the sides 1 to 6, obtained the frequencies 41, 49, 52, 58, 37 and 63 respectively.

(c) Analyse these data and comment on whether or not the assumption is valid for this die. Use a 5% level of significance and state your hypotheses clearly.

(8)

---

6. A sociologist was studying the smoking habits of adults. A random sample of 300 adult smokers from a low income group and an independent random sample of 400 adult smokers from a high income group were asked what their weekly expenditure on tobacco was. The results are summarised below.

	N	mean	s.d.
Low income group	300	£6.40	£6.69
High income group	400	£7.42	£8.13

(a) Using a 5% significance level, test whether or not the two groups differ in the mean amounts spent on tobacco.

(9)

(b) Explain briefly the importance of the central limit theorem in this example.

(2)

---

7. A survey in a college was commissioned to investigate whether or not there was any association between gender and passing a driving test. A group of 50 male and 50 female students were asked whether they passed or failed their driving test at the first attempt. All the students asked had taken the test. The results were as follows.

	Pass	Fail
Male	23	27
Female	32	18

Stating your hypotheses clearly test, at the 10% level, whether or not there is any evidence of an association between gender and passing a driving test at the first attempt.

(11)

---

8. Observations have been made over many years of T, the noon temperature in °C, on 21st March at *Sunnymere*. The records for a random sample of 12 years are given below.

5.2, 3.1, 10.6, 12.4, 4.6, 8.7, 2.5, 15.3, –1.5, 1.8, 13.2, 9.3.

(a) Find unbiased estimates of the mean and variance of T.

(5)

Over the years, the standard deviation of T has been found to be 5.1.

(*b*) Assuming a normal distribution find a 90% confidence interval for the mean of T.

(5)

A meteorologist claims that the mean temperature at noon in *Sunnymere* on 21st March is 4 °C.

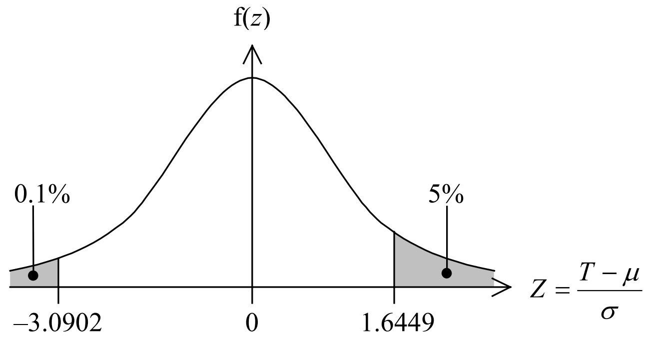
(c) Use your interval to comment on the meteorologist’s claim.

(2)

END

Question Number	Scheme	Marks
1. (*a*)	A statistical process devised to describe or make predictions about the expected behaviour of a real-world problem.	B1 B1 (2)
(*b*) (i)	The number showing on the uppermost side of a die after it has been rolled.	B1
(ii)	The height of adult males.	B1 (2)
		(4 marks)

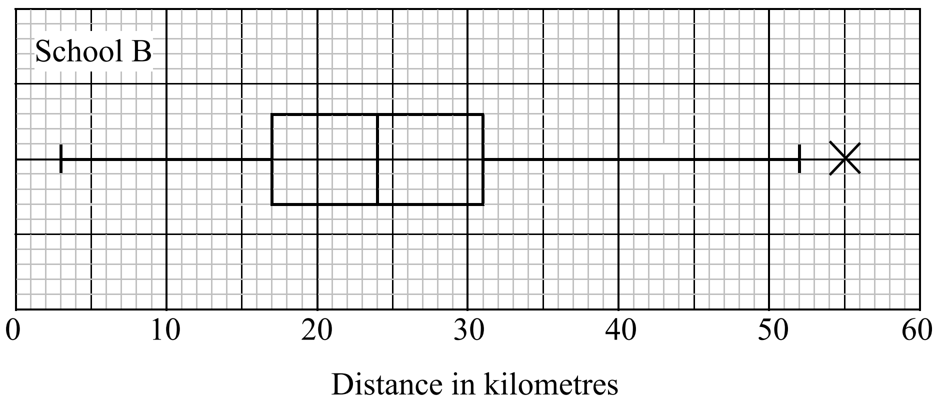
2.



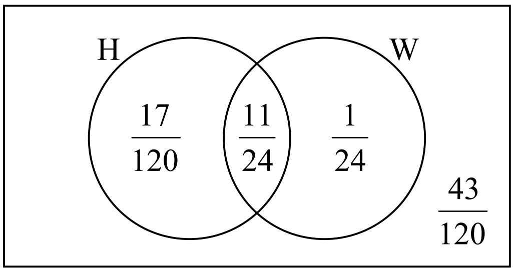
Question Number	Scheme		Marks
2.	$P(T > 55) = 0.05$		
	$\therefore P\left(Z > \dfrac{55 - \mu}{\sigma}\right) = 0.05$	1.6449 Standardising	B1 M1
	$\Rightarrow \dfrac{55-\mu}{\sigma} = 1.6449$	Completely correct	A1
	$P\ (T < 10) = 0.001$		
	$\therefore P\left(Z < \dfrac{10 - \mu}{\sigma}\right) = 0.001$	−3.0902	B1
	$\Rightarrow \dfrac{10-\mu}{\sigma} = -3.0902$	Standardising Completely correct	M1 A1
	$\therefore 55 - \mu = 1.6449\sigma$		
	$10 - \mu = -3.0902\sigma$	Attempt to solve	M1
	$\therefore \mu = 39.368$	$\mu = 39.4$	A1
	$\sigma = 9.5035$	$\sigma = 9.50$	A1 (9)
			(9 marks)

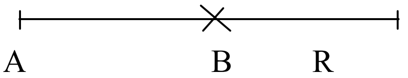
Question Number	Scheme	Marks
3. (*a*)	$k(1+2+3+4+5)=1$ Use of $\sum \mathrm{P}(X=x)=1$	M1 A1
	$\Rightarrow k=\underline{\frac{1}{15}}$ *	A1 (3)
(*b*)	$\mathrm{E}(X)=\frac{1}{15}\{1+2\times 2+\ldots+5\times 5\}$ Use of $\mathrm{E}(X)=\sum x\mathrm{P}(X=x)$	M1 A1
	$=15$	A1
	$\therefore \mathrm{E}(2X+3)=2\mathrm{E}(X)+3$	M1
	$=\frac{31}{3}$	A1 ft (5)
(*c*)	$\mathrm{E}(X^2)=\frac{1}{15}\{1+2^2\times 2+\ldots+5^2\times 5\}$ Use of $\mathrm{E}(X^2)=\sum x^2\mathrm{P}(X=x)$	M1
	$=15$	A1
	$\mathrm{Var}(X)=15-\left(\frac{11}{3}\right)^2$ Use of $\mathrm{Var}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2$	M1
	$=\underline{\frac{14}{9}}$	A1
	$\mathrm{Var}(2X-4)=4\,\mathrm{Var}(X)$ Use of $\mathrm{Var}(aX)=a^2\mathrm{Var}(X)$	M1
	$=\underline{\frac{56}{9}}$	A1 ft (6)
		(14 marks)

Question Number	Scheme		Marks
4. (*a*)	$b = \dfrac{15 \times 484 - 143 \times 391}{15 \times 2413 - (143)^2}$		M1 A1
	$= -3.0899$	AWRT –3.09	A1
	$a = \dfrac{391}{15} - (-3.0899)\left(\dfrac{143}{15}\right)$		M1 A1
	$= 55.5237$	AWRT 55.5	A1
	$\therefore y = 55.52 - 3.09x$		B1 ft
	$\therefore h - 100 = 55.52 - 3.09(s - 20)$		M1 A1 ft
	$\therefore h = 217.32 - 3.09s$	AWRT 217; 3.09	A1 (10)
(*b*)	For every extra revolution/minute the life of the drill is reduced by 3 hours.		B1 B1 (2)
(*c*)	$s = 30 \Rightarrow h = 124.6$	AWRT 125	M1 A1 ft (2)
			(14 marks)

Question Number	Scheme	Marks
5. (*a*)	Advantages: Uses central 50% of the data	
	Not affected by extreme values (outliers)	
	Provide an alternative measure of spread to the variance/standard deviation, i.e. IQR/STQR	
	Disadvantages: Not always a simple calculation, e.g. interpolation for a grouped frequency distribution	
	Different measures of calculation – no single argued method	
	Does not use all the data directly	
	Any 4 sensible comments – at least one advantage and one disadvantage	B1 B1 B1 B1 (4)
(*b*)	Indicates maximum/minimum observations and possible outliers	
	Indicates relative positions of the quartiles	
	Indicates skewness	
	When plotted on the same scale enables comparisons of distributions	
	Any 4 sensible comments	B1 B1 B1 (3)
(*c*)	$Q_1 - 1.5(Q_3 - Q_1) = -4 \Rightarrow$ no outlier below lower quartile	B1
	$Q_2 + 1.5(Q_3 - Q_1) = 52 \Rightarrow$ an outlier (55) above upper quartile	B1
		
	Same scale and label	B1
	Q_1, Q_2, Q_3, 3, 52	B1
	55	B1 (3)
	continued over…	

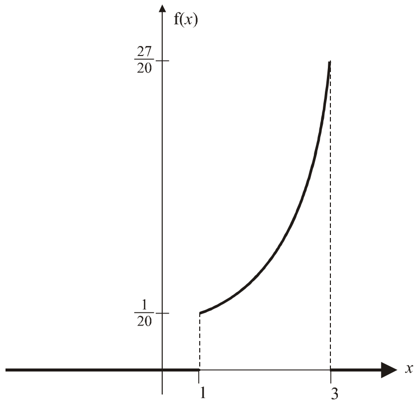
Question Number	Scheme	Marks
5. (*d*)	A: $Q_3 - Q_2 = 10$; $Q_2 - Q_1 = 10 \Rightarrow$ symmetrical B: $Q_3 - Q_2 = 7$; $Q_2 - Q_1 = 7 \Rightarrow$ symmetrical both distributions are symmetrical	
	Median B (24) > Median A (22) $\Rightarrow$ on average teachers in B travel slightly further to school than those in A	
	Range of B is greater than that of A	
	25% of teachers in A travel 12 km or less compared with 25% of teachers in B who travel 17 km or less	
	50% of teachers in A travel between 12 km and 32 km as compared with 17 km and 31 km for B	
	Any 4 sensible comments	B1 B1 B1 B1 (4)
		(16 marks)

Question Number	Scheme		Marks
6. (a)	$\mathrm{P(H \cap W) = P(H \mid W)P(W)}$		M1
	$= \frac{11}{12} \times \frac{1}{2} = \underline{\frac{11}{24}}$ *		A1 (2)
(b)		Diagram	M1
		$\mathrm{H \cap W'}$	M1 A1
		$\mathrm{H' \cap W}$	A1
		$\mathrm{H \cap W}$	B1 (5)
(c)	P (only one has a degree) $= \frac{17}{120} + \frac{1}{24} = \frac{11}{60}$		M1 A1 (2)
(d)	P (neither has a degree) $= 1 - \left\{\frac{17}{120} + \frac{11}{24} + \frac{1}{24}\right\}$		M1 A1
	$= \underline{\frac{43}{120}}$		A1 (3)
(e)	Possibilities	Any one	B1
	$-\mathrm{(HW')(H'W); (H'W)(HW'); (HW)(H'W'); (H'W')(HW)}$		
		All correct	B1
	$\therefore$ P (only 1 H or 1 W) $= \left(2 \times \frac{17}{120} \times \frac{1}{24}\right) + \left(2 \times \frac{11}{24} \times \frac{43}{120}\right)$	$2 \times \frac{17}{120} \times \frac{1}{24}$	B1 ft
	$= \underline{\frac{49}{144}}$	$2 \times \frac{11}{24} \times \frac{43}{120}$	B1 ft
		Adding their probabilities	M1
		$\frac{49}{144}$	A1 (6)
			(18 marks)

Question Number	Scheme	Marks
1. (*a*)	<u>Advantage</u>: eg quicker/cheaper	B1
	<u>Disadvantage</u>: eg doesn't give the full picture	B1 (2)
(*b*)	The register of pupils attending	B1 (1)
(*c*)	The individual pupils	B1 (1)
		(4 marks)
2. (*a*)	 $X \sim \text{U}[0,12]$	B1, B1 (2)
(*b*)	$\text{P}(X \le x) = \int_0^x \frac{1}{12}\,\text{d}t = \underline{\frac{x}{12}}$ $\quad \therefore \text{F}(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{12}, & 0 \le x \le 12 \\ 1 & x > 12 \end{cases}$	M1 A1 B1 ft (centre) B1 (ends) (4)
(*c*)	$\text{P}(X < 4) = \frac{4}{12} = \frac{1}{3}$	B1 ft (1)
		(7 marks)
3. (*a*)	$\text{P(SC)} = \frac{3}{4}$; $\text{P(HC)} = \frac{1}{4}$ either	B1
	Let X represent the number of HC chocolates	
	$\therefore X \sim \text{B}(20; 0.25)$ can be implied	B1
	$\text{P}(X = 10) = 0.9961 - 0.9861 = 0.0100$ awrt 0.010	B1 (3)
(*b*)	$\text{P}(X < 5) = \text{P}(X \le 4)$	M1
	$= 0.4148$ awrt 0.415	A1 (2)
(*c*)	Expected number $= np = 100 \times 0.25 = 25$	M1 A1 (2)
		(7marks)

Question Number		Scheme	Marks	
4.	(*a*)	$\bar{x} = \dfrac{0\times 37 + 1\times 65 + 2\times 60 + \ldots + 5\times 12}{37+65+60+\ldots+12} = \dfrac{500}{250} = 2$	M1 A1cso	(2)
	(*b*)	$\text{var} = \dfrac{\sum x^2}{250} - 2^2 = \dfrac{1478}{250} - 4 = 1.912$ (or $s^2 = 1.9196\ldots$)	M1 A1	(2)
	(*c*)	For a Poisson distribution the mean must equal the variance;		
		parts (a) and (b) are very close, so a Poisson might be a suitable model.	B1	(1)
	(*d*)	H_0: $\mu = 2$; H_1: $\mu < 2$	B1 B1	
		X = number of errors over 4 pages. Under H_0 $X \sim P_0(8)$;	M1	
		$P(X \leq 3) = 0.0424$	M1 A1	
		This is less than 5% so a significant result and there is evidence that the secretary has improved.	A1 ft	(6)
				(11 marks)
5.	(a)	$H_0 : p = 0.30$ $H_1 : p < 30$	B1 B1	
		X = number ordering vegetarian meal $X \sim B(20, 0.30)$ under H_0		
		$P(X \leq 3) = 0.1071 > 5\%$	M1, A1	
		$\therefore$ Not significant i.e. no reason to suspect proportion is lower	A1 ft	(5)
	(b)	$H_0 : p = 0.10$ $H_1 : p \neq 0.10$	B1 B1	
		Y = number ordering vegetarian meal $Y \sim B(100, 0.10) \Rightarrow Y \approx P_0(10)$	M1	
		Need a, b such that $P(Y \leq a) \approx 0.025$ and $P(Y \geq b) \approx 0.025$		
		From tables: $P(Y \leq 4) = 0.0293$ and $P(Y \leq 16) = 0.9730$	M1 A1	
		$\Rightarrow P(Y \geq 17 = 0.0270$	A1	
		$\therefore Y \leq 4$ and $Y \geq 17$		(6)
	(c)	Significance level is $0.0270 + 0.0293 = \underline{0.0563}$ (5.6%)	B1 ft	(1)
				(12 marks)

Question Number		Scheme		Marks	
6.	(a)	X = number of sheep per square	$X \sim \text{P}_0\,(2.25)$	B1	(1)
	(b)	$\text{P}(X=0) = \text{e}^{-2.25} = 0.105399\ldots$	awrt <u>0.105</u>	B1	(1)
	(c)	P ($X > 2$) 1 – P ($X \le 2$), $=1 - \text{e}^{-2.25}\left[1 + 2.25 + \dfrac{(2.25)^2}{2!}\right]$		M1, M1 A1	
		$1 - 0.60933\ldots = 0.39066$	awrt <u>0.391</u>	A1	(4)
	(d)	Sheep would tend to cluster – no longer randomly scattered		B1	(1)
	(e)	$Y \sim \text{P}_0\,(20) \Rightarrow$ normal approx, $\mu = 20$, $\sigma = \sqrt{20}$		M1, A1	
		P ($Y < 15$) = P ($Y \le 14.5$), $= \text{P}\left(Z \le \dfrac{14.5 - 20}{\sqrt{20}}\right)$	$\pm \frac{1}{2}$	M1, M1	
		$= \text{P}\,(Z \le -1.2298\ldots)$		A1	
		$= 1 - 0.8907 = 0.1093$	AWRT <u>0.109</u>	M1 A1	(7)
					(14 marks)

Question Number	Scheme	Marks
7. (*a*)		B1, B1 B1 (3) $\left(\frac{1}{20}, \frac{27}{20}\right)$
(*b*)	$\mathrm{E}(X) = \int_1^3 \frac{1}{20} x^4 \mathrm{d}x = \left[\frac{x^5}{100}\right]_1^3 = \frac{242}{100} = \underline{2.42}$	M1 [M1] A1 (3)
(*c*)	$\sigma^2 = \int_1^3 \frac{1}{20} x^5 \mathrm{d}x - \mu^2 = \left[\frac{x^6}{120}\right]_1^3 - \mu^2 = \frac{728}{120} - (2.42)^2 = 0.21026$	M1 [M1]
	$\therefore \sigma = 0.459$	A1 cso (3)
(*d*)	$\mathrm{P}(X \le x) = \int_1^x \frac{1}{20} t^3 \mathrm{d}t = \left[\frac{t^4}{80}\right]_1^x = \frac{x^4}{80} - \frac{1}{80}$	M1 $[M1]_1^x$ A1 cso
	$\mathrm{F}(x) = \begin{cases} 0 & x \le 1 \\ \frac{1}{80}(x^4 - 1) & 1 < x < 3 \\ 1 & x \ge 3 \end{cases}$	B1 ft, centre B1 ends (5)
(*e*)	$\mathrm{F}(p) = 0.25 \Rightarrow \frac{1}{80}(p^4 - 1) = \frac{1}{4} \therefore p^4 = 21 \Rightarrow p = 2.14...$	M1 A1
	$\mathrm{F}(q) = 0.75 \Rightarrow \frac{1}{80}(q^4 - 1) = \frac{3}{4} \therefore q^4 = 61 \Rightarrow q = 2.79...$	A1
	IQR=<u>0.65</u>	A1 ft (4)
(*f*)	$\mathrm{IQR} \approx \frac{4}{3} \times 0.459 = \underline{0.612}$,	B1
	Sensible comment, e.g. reasonable approximation or slight underestimate	B1 (2)
		(20 marks)

Question Number	Scheme	Marks
1. (*a*)	Label members 1 → 240	B1
	Use random numbers to select first from 1 – 8	B1
	Select every 8^{th} member (e.g. 6,14, 22, …)	B1 (3)
(*b*)	e.g.: More convenient, efficient, faster etc. Any 1	B1 (1)
		(4 marks)
2. (*a*)	$\overline{P} \sim \mathrm{N}\left(110, \frac{8^2}{16}\right)$ ie : $\overline{P} \sim \mathrm{N}\left(110, 2^2\right)$ Normal	B1
	$110, 2^2$	B1 (2)
(*b*)	$\mathrm{P}\left(110 < \overline{P} < 113\right) = \mathrm{P}\left(0 < Z < \frac{113-110}{2}\right)$ Standardising	M1
	$= \mathrm{P}\ (0 < Z < 1.5)$	A1 ft
	$= 0.4332$ AWRT 0.433	A1 (3)
		(5 marks)
3. (a)	Let T represent total time	
	$\therefore \mathrm{E}\ (T) = 225 + 165 + 185 = 575$ 575	B1
	$\mathrm{Var}\ (T) = 38^2 + 23^2 + 27^2 = 2702$ 2702	B1
	$\therefore P\ (533 < T < 655) \quad = P\ (-0.81 < Z < 1.54)$ Standardising	M1 A1 ft
	$= 0.7292$ AWRT 0.729	A1 (5)
(b)	Let D represent the difference in times for tasks B and C (i.e. $B - C$)	
	$\therefore\ \mathrm{E}(D) = 165 - 185 = -20$	B1
	$\mathrm{Var}(D) = 23^2 + 27^2 = 1258$	B1
	$\therefore \mathrm{P}(D > 0) \quad = \mathrm{P}\left(Z > \frac{0-(-20)}{\sqrt{1258}}\right)$ Standardising $-20, \sqrt{1258}$	M1 A1 ft
	$= \mathrm{P}(Z > 0.56)$	
	$= 0.2877$ AWRT 0.288	A1 (5)
		(10 marks)

Question Number	Scheme	Marks
4. (*a*)	Attendance ranks 2, 1, 8, 5, 3, 6, 7, 4	B1
	$\sum d^2 = 48$ Attempt to find $\sum d^2$	M1 A1
	$r_S = 1 - \frac{6\times 48}{8\times 63}$ Substitution of their $\sum d^2$	M1
	$= 0.4286$ awrt 0.429	A1 ft (5)
(*b*)	$H_0 : \rho = 0;\ H_1 : \rho \neq 0.$ both	B1
	With n=8, critical value is 0.7381 0.7381	B1
	Since 0.429 is not in the critical region $(\rho < -0.7381 \text{ or } \rho > 0.7381)$ then there is no evidence to reject H_0 and it can be concluded that at the 5% level there is no evidence of correlation between league position and attendance	
	Correct comparison	M1
	Conclusion	A1 ft (4)
(*c*)	Share ranks evenly.	B1
	Use product moment correlation coefficient on ranks.	B1 (2)
		(11 marks)
5. (*a*)	$P(X = x) = \frac{1}{6};\ x = 1, 2, \ldots, 6.$	B1 B1 (2)
(*b*)	Discrete uniform distribution	B1 (1)
(*c*)	H_0: Discrete uniform distribution is a suitable model	B1
	H_1: Discrete uniform distribution is <u>not</u> a suitable model	B1
	$\alpha = 0.05 \quad \nu = 5;$ CR: $\chi^2 > 11.070$	B1 B1
	$\sum \frac{(O-E)^2}{E} = \frac{1}{50}\{9^2 + 1^2 + 2^2 + 8^2 + 13^2 + 13^2\}$ All E's=50	B1
	$= \frac{448}{50} = \underline{9.76}$ $\sum \frac{(O-E)^2}{E}$	M1 A1
	Since 9.76 is not in the critical region there is no evidence to reject H_0 and thus the data is compatible with the assumption.	A1 ft (8)
		(11 marks)

Question Number	Scheme		Marks
6. (a)	$H_o : \mu_L = \mu_H ; H_1 : \mu_L \neq \mu_H$		B1 B1
	$\text{s.e.} = \sqrt{\frac{8.13^2}{400} + \frac{6.69^2}{300}}$	Substitute into s.e.	M1
		Complete correct expression	A1
	$= 0.5607$	AWRT 0.561	A1
	$\alpha = 0.05 \Rightarrow$ C.R: $z < -1.96$ or $z > 1.96$	± 1.96	B1
	Test statistic: $z = \frac{6.40 - 7.42}{0.5607} = -\underline{1.819}$	$\left(\bar{x}_L - \bar{x}_H\right)$/ their s.e.	M1
		AWRT ± 1.82	A1
	Since -1.819 is not in the critical region then there is no evidence to reject H_0 and thus it can be concluded that there is no difference in mean expenditure on tobacco.		A1 ft (9)
(b)	C. L. Theorem enables use of $\bar{L} \sim \text{Normal}$ and $\bar{H} \sim \text{Normal}$.	$\bar{L}$ or $\bar{H}$	B1
		Normal	B1 (2)
			(11 marks)

Question Number	Scheme		Marks
7.	**Observed Frequencies**		
		Pass / Fail / Total	
	Male: 23, 27, 50		
	Female: 32, 18, 50		
	Total: 55, 45, 100		
	Expected Frequencies		
	Pass / Fail / Total	Use of $\frac{R_T \times C_T}{100}$	M1
	Male: 27.5, 22.5, 50	27.5	A1
	Female: 27.5, 22.5, 50	22.5	A1
	Total: 55, 45, 100		
	H_o : No association between gender and test result		B1
	H_1 : Association between gender and test result		B1
	$\sum\frac{(O-E)^2}{E} = \frac{(23-27.5)^2}{27.5} + \ldots \frac{(18-22.5)^2}{22.5}$	Use of $\sum\frac{(O-E)^2}{E}$	M1 A1
	$= 3.27$		A1
	$\alpha = 0.10 \Rightarrow \chi^2 > 2.705$	$v = 1$	B1
	Since 3.27 is in the critical region there is evidence of association between gender and test result.	2.705	B1
			A1 ft (11)
			(11 marks)

Question Number		Scheme		Marks	
8.	(*a*)	$\bar{x} = \hat{\mu} = \frac{85.2}{12} = \underline{7.10}$		M1A1	
		$s^2 = \frac{1}{11}\left\{906.18 - \frac{(85.2)^2}{12}\right\}$	Substitution in correct formula	M1	
			Complete correct expression	A1 ft	
		$= 27.3873$	AWRT 27.4	A1	(5)
	(*b*)	Confidence interval is given by	$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$	M1	
		$7.10 \pm 1.6449 \times \frac{5.1}{\sqrt{12}}$	Correct expression with their values	A1 ft	
			1.6449	B1	
		ie:- (4.6783, 9.5216)	AWRT (4.68, 9.52)	A1 A1	(5)
	(*c*)	The value 4 is not in the interval;		B1	
		Thus the claim is not substantiated.		B1	(2)
					(12 marks)