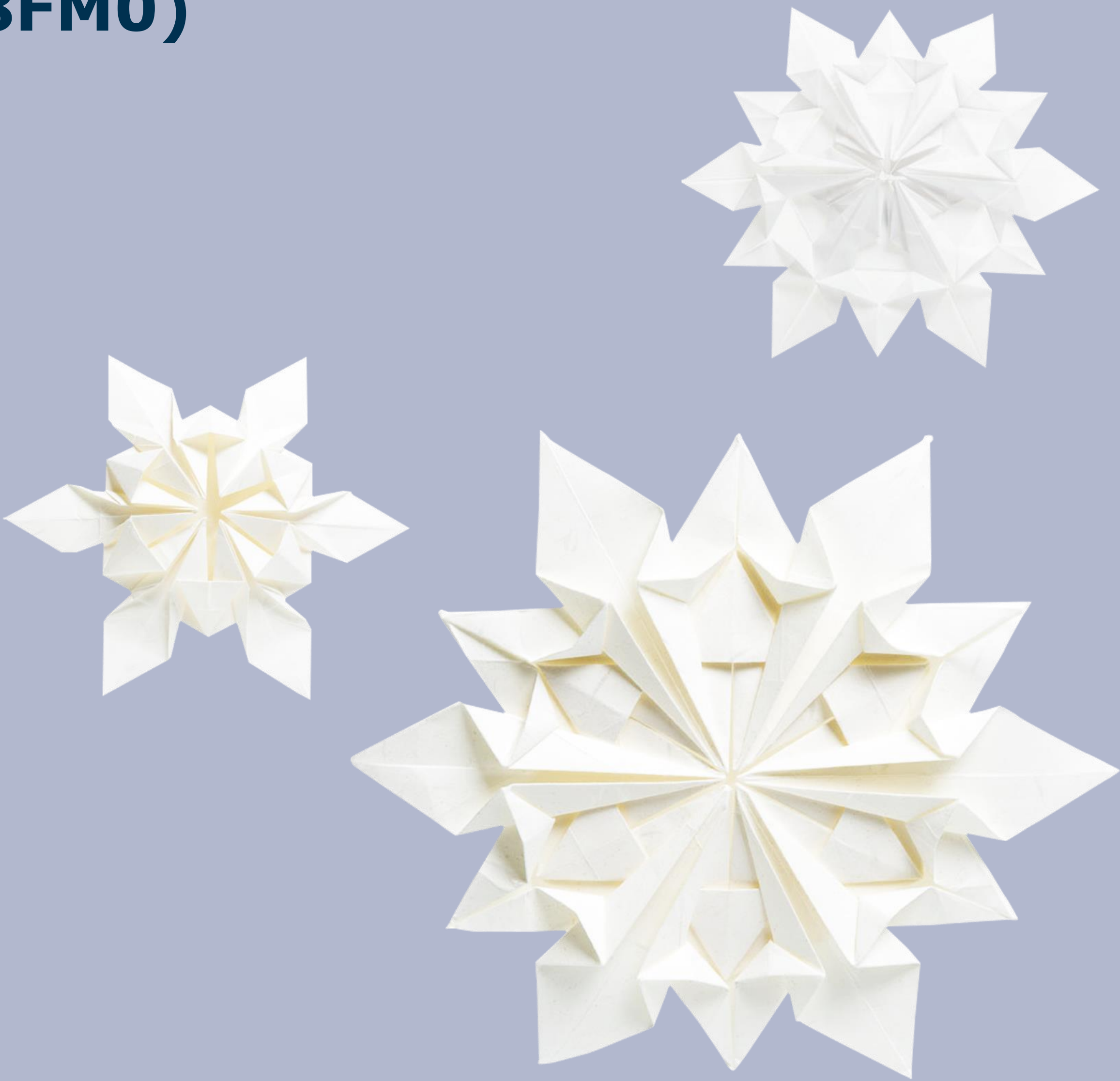
Pearson Edexcel
Level 3 Advanced Subsidiary
GCE in Further Mathematics
(8FM0)



Sample Assessment Materials Model Answers – Further Statistics 1&2

First teaching from September 2017
First certification from June 2018

Sample Assessment Materials Model Answers – Further Statistics 1&2

Contents

Introduction

This booklet has been produced to support mathematics teachers delivering the new Pearson Edexcel Level 3 Advanced Subsidiary GCE in Mathematics (8FMO) specification for first teaching from September 2017.

This booklet looks at Sample Assessment Materials for AS Further Mathematics qualification, specifically at further statistics 1 and 2 questions, and is intended to offer model solutions with different methods explored.

Content of Further Statistics 1&2

Content	**AS level content**
	Further Statistics 1
Discrete probability distribution	Mean and variance of discrete probability distributions. Extension of expected value function to include E(g(X)).
Poisson & binomial distributions	The Poisson distribution. The additive property of Poisson distributions. The mean and variance of the binomial and the Poisson distributions. Use of Poisson distribution as an approximation to the binomial distribution. Extend ideas of hypothesis tests to test for the mean of Poisson distribution.
Chi Squared Tests	Goodness of fit tests and Contingency tables. The null and alternative hypotheses. The use of $\sum_{i=1}^{n}\frac{(O_i - E_i)^2}{E_i}$ as approximate χ^2 statistic. Degrees of freedom.
	Further Statistics 2
Linear Regression	Least squares linear regression. The concept of residuals and minimising the sum of squares of residuals. Residuals. The residual sum of squares (RSS).
Continuous probability distributions	The concept of a continuous random variable. The probability density function and the cumulative distribution function for a continuous random variable. Relationship between probability density and cumulative distribution functions. Mean and variance of continuous random variables. Extension of expected value function to include E(g(X)). Mode, median and percentiles of continuous random variables. Idea of skewness. The continuous uniform (rectangular) distribution.
Correlation	Use of formulae to calculate the product moment correlation. Knowledge of conditions for the use of the product moment correlation. A knowledge of effects of coding. Spearman's rank correlation coefficient, its use and interpretation.

Question 1

A university foreign language department carried out a survey of prospective students to find out which of three languages they were most interested in studying.

A random sample of 150 prospective students gave the following results.

		Language		
		French	**Spanish**	**Mandarin**
Gender	**Male**	23	22	20
	Female	38	32	15

A test is carried out at the 1% level of significance to determine whether or not there is an association between gender and choice of language.

(a) State the null hypothesis for this test.

(1)

H_0: There is no association between language and gender. B1

(b) Show that the expected frequency for females choosing Spanish is 30.6

(1)

$$\frac{(22+32)\times(38+32+15)}{150} = \frac{54\times 85}{150} = 30.6$$ B1

(c) Calculate the test statistic for this test, stating the expected frequencies you have used. **(3)**

Expected frequencies:

Male & French:	$61 \times 65/150 = 26.433$ (3dp)	
Male & Spanish:	$54 \times 65/150 = 23.4$	
Male & Mandarin:	$35 \times 65/150 = 15.167$ (3dp)	
Female & French:	$61 \times 85/150 = 34.567$ (3dp)	
Female & Mandarin:	$35 \times 85/150 = 19.833$ (3dp)	M1

$$\chi^2 = \sum\frac{(O-E)^2}{E} = \frac{(23-26.433)^2}{26.433} + \frac{(22-23.4)^2}{23.4} + \frac{(20-15.167)^2}{15.167} + \frac{(38-34.567)^2}{34.567}$$

$$+\frac{(32-30.6)^2}{30.6} + \frac{(15-19.833)^2}{19.833}$$ M1

$= 0.446 + 0.084 + 1.540 + 0.341 + 0.064 + 1.178$

$= 3.65$ (2dp) A1

(d) State whether or not the null hypothesis is rejected. Justify your answer. **(2)**

Formula book: Percentage points of the χ^2 distribution

Degrees of freedom, $v = (3-1)\times(2-1) = 2$

1% level of significance: 0.010 column.

Critical value: $\chi^2 = 9.210$ M1

As $3.65 < 9.210$, the null hypothesis is not rejected. A1

(e) Explain whether or not the null hypothesis would be rejected if the test was carried out at the 10% level of significance. **(1)**

10% level of significance: 0.100 column.

Critical value: $\chi^2 = 4.605$

The null hypothesis is still not rejected since $3.65 < 4.605$ B1

Question 2

The discrete random variable X has probability distribution given by

x	-1	0	1	2	3
$P(X = x)$	c	a	a	b	c

The random variable $Y = 2 - 5X$
Given that $E(Y) = -4$ and $P(Y \geq -3) = 0.45$

(a) find the probability distribution of X.

(7)

$E(Y) = 2 - 5E(X)$
$-4 = 2 - 5E(X)$ M1
$E(X) = 1.2$
$-1\times c + 0\times a + 1\times a + 2\times b + 3\times c = 1.2$ M1
$a + 2b + 2c = 1.2$ ---[1]

$P(Y \geq -3) = 0.45$
$P(2 - 5X \geq -3) = 0.45$
$P(X \leq 1) = 0.45$
$2a + c = 0.45$ ---[2] M1

sum of probabilities = 1
$2a + b + 2c = 1$ ---[3] M1

Use elimination to solve the simultaneous equations.

[3] – [2]: $b + c = 0.55$
sub into [1]: $a + 2(b + c) = 1.2$
$a + 2(0.55) = 1.2$
$a = 0.1$ M1
$b = 0.3$ A1
$c = 0.25$ A1

Alternative: (for final 3 marks)

Put these 3 simultaneous equations into a matrix equation.

$$\begin{pmatrix} 1 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1.2 \\ 0.45 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 2 \end{pmatrix}^{-1}\begin{pmatrix} 1.2 \\ 0.45 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 2 & -3 \\ -2 & -3 & 4 \end{pmatrix}\begin{pmatrix} 1.2 \\ 0.45 \\ 1 \end{pmatrix}$$ M1

$a = 0.1$
$b = 0.3$ A1
$c = 0.25$ A1

Note: Candidates are expected to have a suitable calculator that will perform calculations with matrices up to at least a 3×3 matrix.

Given also that $\mathrm{E}(Y^2) = 75$ (b) find the exact value of Var (X) **(2)**

$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$
$= 75 - (-4)^2$ M1
$= 59$
$\mathrm{Var}(Y) = 5^2\,\mathrm{Var}(X)$
$\mathrm{Var}(X) = \dfrac{59}{25} = 2.36$ A1

(c) Find $\mathrm{P}(Y > X)$ **(2)**

$\mathrm{P}(Y > X) = \mathrm{P}(2 - 5X > X) = \mathrm{P}(X < \tfrac{1}{3})$ M1

$\mathrm{P}(X < \dfrac{1}{3}) = a + c$
$= 0.35$ A1

Question 3

Two car hire companies hire cars independently of each other.

Car Hire A hires cars at a rate of 2.6 cars per hour.

Car Hire B hires cars at a rate of 1.2 cars per hour.

(a) In a 1 hour period, find the probability that each company hires exactly 2 cars.

(2)

Let $A \sim \text{Po}(2.6)$, $B \sim \text{Po}(1.2)$

P(each hire 2 in 1 hour) = P(A = 2) × P(B = 2)
= 0.25104... × 0.21685... M1
= 0.0544 (3sf) A1

(See the note about suitable calculators at the end of this question)

(b) In a 1 hour period, find the probability that the total number of cars hired by the two companies is 3.

(2)

Let $W = A + B$, then $W \sim \text{Po}(3.8)$ M1

P(W = 3) = 0.20458... = 0.205 (3sf) A1

(c) In a 2 hour period, find the probability that the total number of cars hired by the two companies is less than 9.

(2)

(2.6+1.2)×2 = 7.6, so let $T \sim \text{Po}(7.6)$ M1

P(T < 9) = 0.64819... = 0.648 (3sf) A1

On average, 1 in 250 new cars produced at a factory has a defect.

In a random sample of 600 new cars produced at the factory,

(d) (i) find the mean of the number of cars with a defect,

(ii) find the variance of the number of cars with a defect.

(2)

From the formula book:
Binomial distribution: Mean = np, Variance = $np(1 - p)$

Binomial, $n = 600$, $p = 0.004$, i.e. $X \sim B(600, 0.004)$
(i)
Mean = $600 \times 0.004 = 2.4$ B1
(ii)
Variance = $600 \times 0.004 \times (1 - 0.004) = 2.3904 = 2.39$ (3sf) B1

(e) (i) Use a Poisson approximation to find the probability that no more than 4 of the cars in the sample have a defect.

(ii) Give a reason to support the use of a Poisson approximation.

(2)

(i)
Let D ~ Po(2.4)
$P(D \le 4) = 0.9041... = 0.904$ (3sf) B1

(ii)
Since n is large and p is small.
or
Since mean is approximately equal to variance. B1

<u>Note:</u> Candidates are expected to have a suitable calculator that can access probabilities from standard statistical distributions.

While parts (a) and (b) can be easily done using the formula for the Poisson distribution: $P(X = x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$, available in the formula book, it would be rather time consuming to do parts (c) and (e) using this formula.

Question 4

The discrete random variable X follows a Poisson distribution with mean 1.4

(a) Write down the value of

(i) $P(X = 1)$

(ii) $P(X \leq 4)$

(2)

(i)
$P(X = 1) = 0.34523... = 0.345$ (3sf) B1

(ii)
$P(X \leq 4) = 0.98575... = 0.986$ (3sf) B1

(See the note about suitable calculators at the end of Q3)

The manager of a bank recorded the number of mortgages approved each week over a 40 week period.

Number of mortgages approved	0	1	2	3	4	5	6
Frequency	10	16	7	4	2	0	1

(b) Show that the mean number of mortgages approved over the 40 week period is 1.4

(1)

$$\frac{(0\times10)+(1\times16)+(2\times7)+(3\times4)+(4\times2)+(5\times0)+(6\times1)}{40}=\frac{56}{40}=1.4$$ B1

The bank manager believes that the Poisson distribution may be a good model for the number of mortgages approved each week.

She uses a Poisson distribution with a mean of 1.4 to calculate expected frequencies as follows.

Number of mortgages approved	0	1	2	3	4	5 or more
Expected frequency	9.86	r	9.67	4.51	1.58	s

(c) Find the value of r and the value of s, giving your answers to 2 decimal places.

(2)

$r = 40 \times 0.34523... = 13.81$ M1

$s = 40 \times (1 - 0.98575...) = 0.57$ A1

The bank manager will test, at the 5% level of significance, whether or not the data can be modelled by a Poisson distribution.

(d) Calculate the test statistic and state the conclusion for this test. State clearly the degrees of freedom and the hypotheses used in the test.

(6)

H_0: The Poisson distribution is a suitable model
H_1: The Poisson distribution is not a suitable model B1
(λ is not defined in the hypothesis)

Cells need to be combined when expected frequencies < 5
so combine the last 3 cells: 4.51 + 1.58 + 0.57 = 6.66
so combine the last 4 observed frequencies: 4 + 2 + 0 + 1 = 7 M1

$$\chi^2 = \sum\frac{(O-E)^2}{E} = \frac{(10-9.86)^2}{9.86} + \frac{(16-13.81)^2}{13.81} + \frac{(7-9.67)^2}{9.67} + \frac{(7-6.66)^2}{6.66}$$ M1

$= 0.002 + 0.347 + 0.737 + 0.017$
$= 1.10$ (3sf) A1

Degrees of freedom, $v = 4 - 1 - 1 = 2$ B1
(The data was used to find 1 parameter and the totals must agree)

Formula book: Percentage points of the χ^2 distribution
5% level of significance: 0.05 column.
Critical value: $\chi^2 = 5.991$
Do not reject H_0 since $1.10 < 5.991$
The number of mortgages approved each week follows a Poisson distribution. A1

Question 5

In a gymnastics competition, two judges scored each of 8 competitors on the vault.

Competitor	A	B	C	D	E	F	G	H
Judge 1's scores	4.6	9.1	8.4	8.8	9.0	9.5	9.2	9.4
Judge 2's scores	7.8	8.8	8.6	8.5	9.1	9.6	9.0	9.3

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

(4)

From the formula book:

Spearman's rank correlation coefficient: $r_s = 1 - \dfrac{6\sum d^2}{n(n^2-1)}$

Competitor	A	B	C	D	E	F	G	H
Judge 1's ranks	8	4	7	6	5	1	3	2
Judge 2's ranks	8	5	6	7	3	1	4	2
d^2	0	1	1	1	4	0	1	0

(ranks) M1
(d^2) M1

$\Sigma d^2 = 8$

$r_s = 1 - \dfrac{6 \times 8}{8(64-1)}$ dM1

r_s = 0.90476... = 0.905 (3sf) A1

(b) Stating your hypotheses clearly, test at the 1% level of significance, whether or not the two judges are generally in agreement.

(4)

H_0: $\rho_s = 0$ H_1: $\rho_s > 0$ B1

Formula book: Critical values for correlation coefficients
n = 8, level: 1% = 0.01
critical value ρ_s = 0.8333 B1

r_s = 0.905 > 0.8333 so reject H_0, r_s lies in the critical region. M1
The two judges are in agreement. A1

(c) Give a reason to support the use of Spearman's rank correlation coefficient in this case.

(1)

The data is unlikely to be from a bivariate normal distribution.
or
The emphasis here is on the ranks and not the individual scores. B1

The judges also scored the competitors on the beam.

Spearman's rank correlation coefficient for their ranks on the beam was found to be 0.952

(d) Compare the judges' ranks on the vault with their ranks on the beam.

(1)

Both show positive correlation, but the judges agree more on the beam, since 0.952 is closer to 1. B1

Question 6

The continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{18}(11-2x) & 1 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find P(X < 3)

(2)

$$P(X < 3) = \int_1^3 \frac{1}{18}(11-2x)\mathrm{d}x$$ M1

$$= \left[\frac{1}{18}\left(11x - x^2\right)\right]_1^3$$

$$= \frac{1}{18}\ ((33-9)-(11-1))$$

$$= \frac{7}{9}$$ A1

<u>Alternative:</u>

Area of Trapezium $= \frac{1}{2}\left(\frac{9}{18}+\frac{5}{18}\right)\times 2$ M1

$= \frac{7}{9}$ A1

(b) State, giving a reason, whether the upper quartile of X is greater than 3, less than 3 or equal to 3.

(1)

Since P(X < 3) > 0.75, the upper quartile is less than 3. B1

Given that $\mathrm{E}(X) = \frac{9}{4}$

(c) use algebraic integration to find Var(X)

(3)

$$\mathrm{E}(X^2) = \int_1^4 \frac{1}{18}x^2(11-2x)\,\mathrm{d}x \qquad \text{M1}$$

$$= \left[\frac{1}{18}\left(\frac{11}{3}x^3 - \frac{1}{2}x^4\right)\right]_1^4$$

$$= \frac{1}{18}\left(\left(\frac{704}{3}-128\right)-\left(\frac{11}{3}-\frac{1}{2}\right)\right)$$

$$= \frac{23}{4}$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

$$= \frac{23}{4} - \left(\frac{9}{4}\right)^2 \qquad \text{M1}$$

$$= \frac{11}{16} \qquad \text{A1}$$

The cumulative distribution function of X is given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{18}(11x - x^2 + c) & 1 \le x \le 4 \\ 1 & x > 4 \end{cases}$$

(d) Show that $c = -10$.

(2)

Either:	**Or:**	
$\mathrm{F}(4) = 1$	$\mathrm{F}(1) = 0$	
$\frac{1}{18}(11(4) - 4^2 + c) = 1$	$\frac{1}{18}(11(1) - 1^2 + c) = 0$	M1
$28 + c = 18$	$10 + c = 0$	
$c = -10$	$c = -10$	A1

(e) Find the median of X, giving your answer to 3 significant figures. **(3)**

$F(m) = 0.5$	M1
$\frac{1}{18}(11m - m^2 - 10) = 0.5$	
$m^2 - 11m + 19 = 0$	
$m = \frac{-(-11) \pm \sqrt{(-11)^2 - 4(1)(19)}}{2(1)}$	M1
$m = 2.1458...$ or $8.8541...$	
reject 8.85 as it is not in the range 1 to 4.	
$m = 2.15$	A1

Question 7

A scientist wants to develop a model to describe the relationship between the average daily temperature, x °C, and a household's daily energy consumption, y kWh, in winter.

A random sample of the average temperature and energy consumption are taken from 10 winter days and are summarised below.

$$\sum x = 12 \qquad \sum x^2 = 24.76 \qquad \sum y = 251 \qquad \sum y^2 = 6341 \qquad \sum xy = 284.8$$

$$S_{xx} = 10.36 \qquad S_{yy} = 40.9$$

(a) Find the product moment correlation coefficient between y and x.

(2)

From the formula book:

For a set of n pairs of values (x_i, y_i): $S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$

Product moment correlation coefficient: $r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$

$S_{xy} = 284.8 - \frac{12 \times 251}{10} = -16.4$ M1

$r = \frac{-16.4}{\sqrt{10.36 \times 40.9}}$

$r = -0.79671... = -0.797$ (3sf) A1

(b) Find the equation of the regression line of y on x in the form $y = a + bx$.

(3)

$b = \frac{-16.4}{10.36} = -1.583$ M1

$\bar{x} = \frac{12}{10} = 1.2, \ \bar{y} = \frac{251}{10} = 25.1$

$a = 25.1 - (-1.583...)1.2 = 26.99...$ M1

$y = 27.0 - 1.58x$ A1

(c) Use your equation to estimate the daily energy consumption when the average daily temperature is 2 °C

(1)

$y = 27.0 - 1.58\times2 = 23.84$ B1

(d) Calculate the residual sum of squares (RSS).

(2)

From the formula book:

Residual Sum of Squares (RSS) = $S_{yy} - \dfrac{(S_{xy})^2}{S_{xx}} = S_{yy}(1 - r^2)$

Either:

RSS = $40.9 - \dfrac{-16.4}{10.36}$ M1

= 14.938... = 14.9 (3sf) A1

Or:

RSS = $40.9(1 - (-0.79671)^2)$ M1

= 14.938... = 14.9 (3sf) A1

The table shows the residual for each value of x.

x	–0.4	–0.2	0.3	0.8	1.1	1.4	1.8	2.1	2.5	2.6
Residual	–0.63	–0.32	–0.52	–0.73	0.74	2.22	1.84	0.32	f	–1.88

(e) Find the value of f.

(2)

Sum of residuals = 0

$-0.63 - 0.32 - 0.52 - 0.73 + 0.74 + 2.22 + 1.84 + 0.32 + f - 1.88 = 0$ M1

$f = -1.04$ A1

(f) By considering the signs of the residuals, explain whether or not the linear regression model is a suitable model for these data.

(1)

The residuals should be randomly scattered above and below zero so a linear model may not be appropriate. B1

Question 8

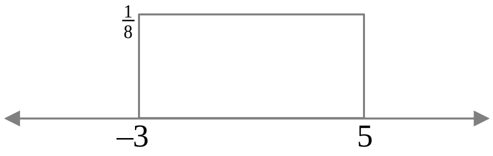
The continuous random variable X is uniformly distributed over the interval [–3, 5].

(a) Sketch the probability density function f(x) of X.

(2)

Uniform distribution has a rectangular shape. B1

width of rectangle = 5 – (–3) = 8 so height = $\frac{1}{8}$ since area = 1.



B1

(b) Find the value of k such that P(X < 2[k – X]) = 0.25

(3)

$P(X < 2(k - X)) = P(X < \frac{2}{3}k)$ M1

$$\frac{\frac{2}{3}k - (-3)}{5 - (-3)} = 0.25$$ M1

$\frac{2}{3}k + 3 = 2$

$k = -1.5$ A1

<u>Alternative:</u>

$P(X < -1) = 0.25$ so $2(k - x) = -1$ and $x = -1$ M1

$2(k - (-1)) = -1$

$k + 1 = -0.5$ M1

$k = -1.5$ A1

(c) Use algebraic integration to show that $E(X^3) = 17$

(3)

$E(X^3) = \int_{-3}^{5} \frac{1}{8} x^3 \mathrm{d}x$ M1

$= \left[\frac{1}{32} x^4\right]_{-3}^{5}$

$= \frac{1}{32} (5^4 - (-3)^4)$ dM1

$= 17$ A1

For more information on Edexcel and BTEC qualifications please visit our websites:
www.edexcel.com and www.btec.co.uk

Pearson Education Limited. Registered in England and Wales No. 872828
Registered Office: 80 Strand, London WC2R 0RL.
VAT Reg No GB 278 537121