GCE AS/A LEVEL – NEW

2305U20-1

S19-2305U20-1

THURSDAY, 16 MAY 2019 – AFTERNOON

FURTHER MATHEMATICS – AS unit 2
FURTHER STATISTICS A

1 hour 30 minutes

2305U201
01

ADDITIONAL MATERIALS

In addition to this examination paper, you will need:

- a WJEC pink 16-page answer booklet;
- a Formula Booklet;
- a calculator;
- statistical tables (RND/WJEC Publications).

INSTRUCTIONS TO CANDIDATES

Use black ink or black ball-point pen.

Answer **all** questions.

Sufficient working must be shown to demonstrate the **mathematical** method employed.

Answers without working may not gain full credit.

Unless the degree of accuracy is stated in the question, answers should be rounded appropriately.

INFORMATION FOR CANDIDATES

The maximum mark for this paper is 70.

The number of marks is given in brackets at the end of each question or part-question.

You are reminded of the necessity for good English and orderly presentation in your answers.

© WJEC CBAC Ltd.

stemup.co.uk

Reminder: *Sufficient working must be shown to demonstrate the **mathematical** method employed.*

1. *(a)* Sketch a scatter diagram of a dataset for which Spearman's rank correlation coefficient is +1, but the product moment correlation coefficient is less than 1. [2]

Two judges were judging cheese at the UK Cheese Festival. There were 8 blue cheeses in a particular category. The rankings are shown below.

Cheese	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Judge 1	1	5	8	7	6	4	3	2
Judge 2	1	3	8	5	2	4	6	7

(b) Calculate Spearman's rank correlation coefficient for this dataset. [3]

(c) By sketching a scatter diagram of the rankings, or otherwise, comment on the extent to which the judges agree. [2]

2. The probability of winning a certain game at a funfair is p. Aman plays the game 5 times and Boaz plays the game 8 times. The independent random variables X and Y denote the number of wins for Aman and Boaz respectively.

(a) Given that E(XY) = 6·4, calculate p. [3]

(b) Find Var(XY). [6]

3. The number of claims made to the home insurance department of an insurance company follows a Poisson distribution with mean 4 per day.

(a) Find the probability that more than 11 claims are made in a 2-day period. [3]

The number of claims made in a day to the pet insurance department of the same company follows a Poisson distribution with parameter λ. An insurance company worker notices that the probability of two claims being made in a day is three times the probability of four claims being made in a day.

(b) Determine the value of λ. [3]

The car insurance department models the length of time between claims for drivers aged 17 to 21 as an exponential distribution with mean 10 months. Rachel is 17 years old and has just passed her test. Her father says he will give her the car that they share if she does not make a claim in the first 12 months.

(c) What is the probability that her father gives her the car? [3]

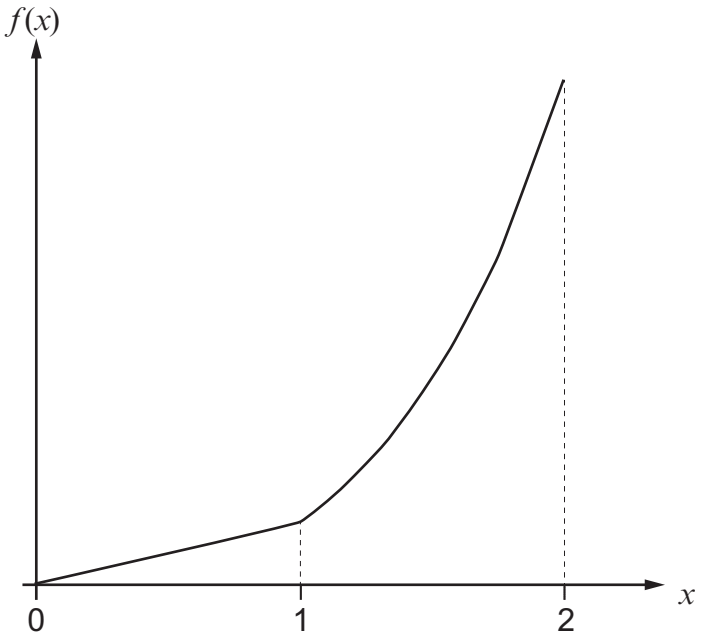
© WJEC CBAC Ltd.

stemup.co.uk

4. The continuous random variable, X, has the following probability density function

$$f(x) = \begin{cases} kx & \text{for} \quad 0 \leqslant x < 1, \\ kx^3 & \text{for} \quad 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.



(a) Show that $k = \frac{4}{17}$. [4]

(b) Determine E(X). [4]

(c) Calculate E($3X - 1$) and Var($3X - 1$). [7]

TURN OVER

© WJEC CBAC Ltd.

stemup.co.uk

5. Chris is investigating the distribution of birth months for ice hockey players. He collects data for 869 randomly chosen National Hockey League (NHL) players. He decides to carry out a chi-squared test. Using a spreadsheet, he produces the following output.

	A	B	C	D
1	**Birth Month**	**Observed**	**Expected**	**Chi-Squared Contributions**
2	Jan-Mar	259	217.25	8.023302647
3	Apr-June	232	217.25	1.001438435
4	Jul-Sept	200	217.25	1.369677791
5	Oct-Dec	178	217.25	7.091196778
6	Total	869	869	17.48561565
7				
8			p value	
9			0.000561458	
10				

(a) By considering the output, state the null hypothesis that Chris is testing. State what conclusion Chris should reach and explain why. [3]

Chris now wonders if Premier League football players' birth months are distributed uniformly throughout the year. He collects the birth months of 75 randomly selected Premier League footballers. This information is shown in the table below.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	7	11	4	12	2	6	6	5	8	5	6

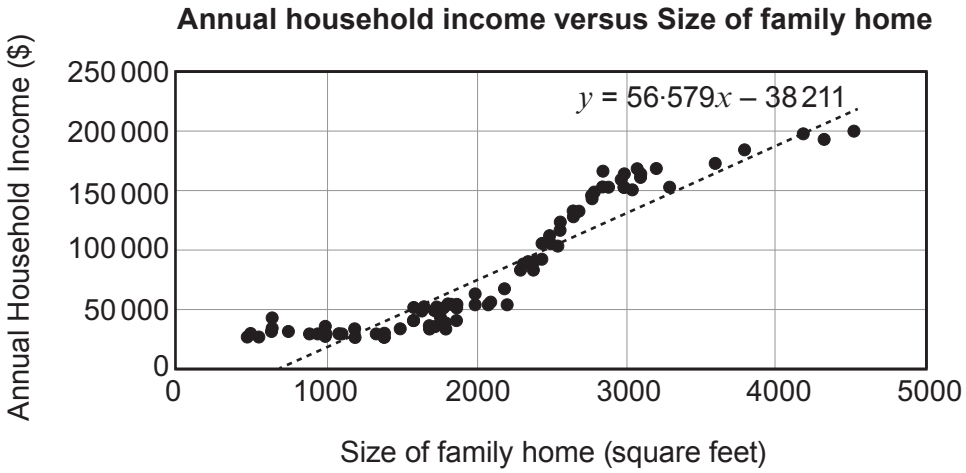
(b) Carry out the chi-squared goodness of fit test at the 10% significance level that Chris should use to conduct his investigation. [8]

© WJEC CBAC Ltd.

stemup.co.uk

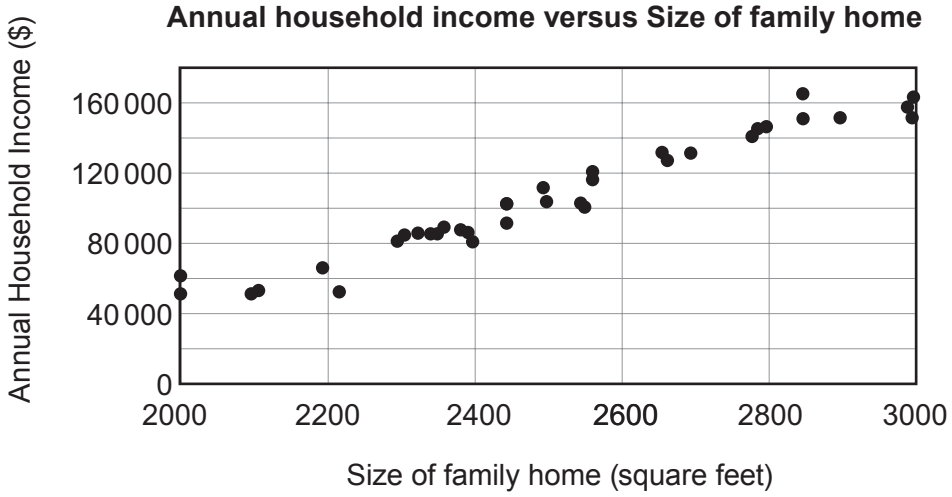
6. The University of Arizona surveyed a large number of households. One purpose of the survey was to determine if annual household income could be predicted from size of family home.

The graph of *Annual household income*, y, versus *Size of family home*, x, is shown below.



(a) State the limitations of using the regression line above with reference to the scatter diagram. [1]

The data for size of family homes between 2000 and 3000 square feet are shown in the diagram below.



Summary statistics for these data are as follows.

$$\sum x = 93160 \qquad \sum y = 3\,907\,142 \qquad n = 37$$

$$S_{xx} = 2\,869\,673{\cdot}03 \qquad S_{yy} = 44\,312\,797\,167 \qquad S_{xy} = 348\,512\,820{\cdot}6$$

(b) Calculate the equation of the least squares regression line to predict *Annual household income* from *Size of family home* for these data. [5]

TURN OVER

© WJEC CBAC Ltd.

stemup.co.uk

7. An article published in a medical journal investigated sports injuries in adolescents' ball games: football, handball and basketball. In a study of 906 randomly selected adolescent players in the three ball games, 379 players incurred injuries over the course of one year of playing the sport. Rhian wants to test whether there is an association between the site of injury and the sport played. A summary of the injuries is shown in the table below.

		Site of injury							
	Observed values	**Shoulder/ Arm**	**Hand/ Fingers**	**Thigh/ Leg**	**Knee**	**Ankle**	**Foot**	**Other**	**Total**
Sport	**Football**	8	3	45	36	51	36	12	191
	Handball	14	26	6	15	42	6	6	115
	Basketball	4	28	4	4	22	1	10	73
	Total	26	57	55	55	115	43	28	379

(a) Calculate the values of *A*, *B*, *C* in the tables below. [4]

		Site of injury						
	Expected values	**Shoulder/ Arm**	**Hand/ Fingers**	**Thigh/ Leg**	**Knee**	**Ankle**	**Foot**	**Other**
Sport	**Football**	13·1029	28·7256	27·7177	27·7177	57·9551	21·6702	14·1108
	Handball	7·8892	17·2955	16·6887	16·6887	*A*	13·0475	8·4960
	Basketball	5·0079	10·9789	10·5937	10·5937	22·1504	8·2823	5·3931

		Site of injury						
	Chi-Squared Contributions	**Shoulder/ Arm**	**Hand/ Fingers**	**Thigh/ Leg**	**Knee**	**Ankle**	**Foot**	**Other**
Sport	**Football**	1·98732	23·03890	10·77575	2·47484	*B*	9·47586	0·31575
	Handball	4·73333	4·38079	*C*	0·17087	1·44690	3·80664	0·73331
	Basketball	0·20286	26·38865	4·10400	4·10400	0·00102	6·40306	3·93521

(b) Given that the test statistic, X^2, is 116·16, carry out the significance test at the 5% level. [5]

(c) Which site of injury most affects the conclusion of this test? Comment on your answer. [2]

© WJEC CBAC Ltd.

stemup.co.uk

Rhian also analyses the data on the type of contact that caused the injuries and the sport in which they occur, shown in the table below.

Observed values	**Ball**	**Opponent**	**Surface**	**None**	**Total**
Football	17	68	17	92	194
Handball	23	34	19	38	114
Basketball	28	17	12	14	71
Total	68	119	48	144	379

The chi-squared test statistic is 46·0937. Rhian notes that this value is smaller than 116·16, the test statistic in part *(b)*. She concludes that there is weaker evidence for association in this case than there was in part *(b)*.

(d) State Rhian's misconception and explain what she should consider instead. [2]

END OF PAPER

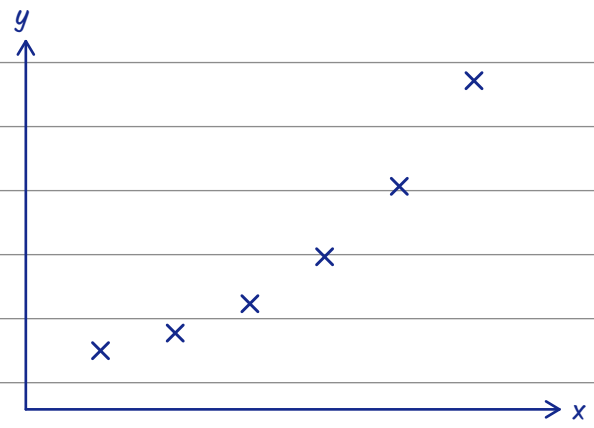
© WJEC CBAC Ltd.

stemup.co.uk

Surname	StemUp Model Solution
Answer booklet	Write the question number in the left-hand margin at the start of each answer.

Question number

1 (a)



Ranks of x and y increase together, so $r_s = +1$, but the points do not lie on a straight line, so $r < 1$.

1 (b)

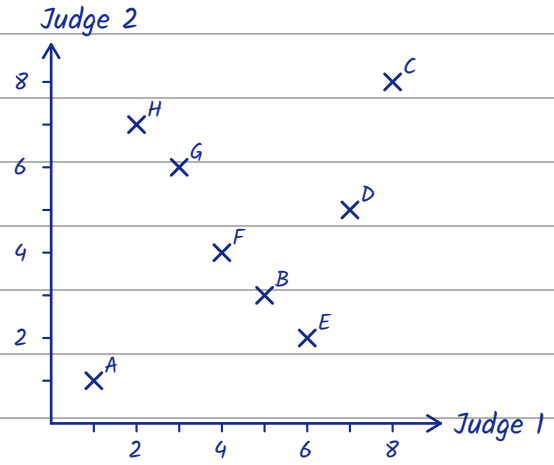
$d = 0, 2, 0, 2, 4, 0, -3, -5$

$d^2 = 0, 4, 0, 4, 16, 0, 9, 25 \qquad \Sigma d^2 = 58$

$r_s = 1 - 6\Sigma d^2 \div (n(n^2 - 1)) = 1 - (6 \times 58) \div (8 \times 63)$

$r_s = 1 - 348 \div 504 = 0.3095\ldots = 13/42$

1 (c)



The judges agree on the best cheese (A) and the worst cheese (C), and they agree exactly on 3 of the 8 cheeses (A, C, F).

But they almost completely disagree about the others (e.g. E and H), so overall there is only weak agreement ($r_s = 0.31$).

Question number

2 (a)

$X \sim B(5, p)$ $Y \sim B(8, p)$

X and Y independent, so $E(XY) = E(X) \times E(Y)$

$E(XY) = 5p \times 8p = 40p^2 = 6.4$

$p^2 = 0.16 \Rightarrow p = 0.4 \quad (p > 0)$

2 (b)

$Var(X) = 5 \times 0.4 \times 0.6 = 1.2$ $Var(Y) = 8 \times 0.4 \times 0.6 = 1.92$

$E(X^2) = Var(X) + (E(X))^2 = 1.2 + (0.4 \times 5)^2 = 5.2$

$E(Y^2) = Var(Y) + (E(Y))^2 = 1.92 + (0.4 \times 8)^2 = 12.16$

$Var(XY) = E(X^2Y^2) - (E(XY))^2 = E(X^2)E(Y^2) - (E(X)E(Y))^2$

$Var(XY) = 5.2 \times 12.16 - 6.4^2 = 63.232 - 40.96$

$Var(XY) = 22.272$

3 (a)

Let X = number of claims in a 2-day period, $X \sim Po(8)$

$P(X > 11) = 1 - P(X \leq 11) = 1 - 0.8881$

$P(X > 11) = 0.1119$

3 (b)

Let Y = number of pet insurance claims in a day, $Y \sim Po(\lambda)$

$P(Y = 2) = 3 \times P(Y = 4)$

$\lambda^2 e^{-\lambda} \div 2! = 3 \times \lambda^4 e^{-\lambda} \div 4!$

$\lambda^2 \div 2 = \lambda^4 \div 8 \Rightarrow 24\lambda^2 = 6\lambda^4$

$\lambda^2(\lambda^2 - 4) = 0 \Rightarrow \lambda = 0, 2$ or -2

$\lambda > 0$ ($\lambda = 0$ would give zero probabilities), so $\lambda = 2$

3 (c)

T = time (months) until her first claim, $T \sim$ Exp with mean 10, so $\lambda = 1/10$

$P(T > 12) = \int_{12}^{\infty} (1/10)e^{-t/10}\, dt = [-e^{-t/10}]_{12}^{\infty}$

$P(T > 12) = e^{-12/10} = e^{-1.2}$

$P(T > 12) = 0.30119\ldots = 0.301$

Question number	
4 (a)	$\int f(x)\,dx = 1$: $\int_0^1 kx\,dx + \int_1^2 kx^3\,dx = 1$
	$[kx^2/2]_0^1 + [kx^4/4]_1^2 = 1$
	$k/2 + (4k - k/4) = 1$
	$k/2 + 15k/4 = 1 \Rightarrow 17k/4 = 1$
	$\therefore k = 4/17$ (as required)
4 (b)	$E(X) = \int x\,f(x)\,dx = \int_0^1 kx^2\,dx + \int_1^2 kx^4\,dx$
	$E(X) = [kx^3/3]_0^1 + [kx^5/5]_1^2$
	$E(X) = k/3 + (32k/5 - k/5) = k/3 + 31k/5 = 98k/15$
	$E(X) = (98/15) \times (4/17) = 392/255 = 1.5372\ldots = 1.54$ (3 s.f.)
4 (c)	$E(3X - 1) = 3E(X) - 1 = 3 \times 392/255 - 1$
	$E(3X - 1) = 307/85 = 3.6117\ldots = 3.61$
	$E(X^2) = \int_0^1 kx^3\,dx + \int_1^2 kx^5\,dx = [kx^4/4]_0^1 + [kx^6/6]_1^2$
	$E(X^2) = k/4 + (64k/6 - k/6) = k/4 + 21k/2 = 43k/4 = 43/17$
	$\text{Var}(X) = E(X^2) - (E(X))^2 = 43/17 - (392/255)^2$
	$\text{Var}(X) = 2.52941\ldots - 2.36315\ldots = 0.16625\ldots = 0.166$
	$\text{Var}(3X - 1) = 3^2\,\text{Var}(X) = 9 \times 0.16625\ldots$
	$\text{Var}(3X - 1) = 1.4963\ldots = 1.496$
5 (a)	H_0: birth months of NHL players are evenly (uniformly) distributed across the year, i.e. each quarter is equally likely.
	The p-value 0.000561 is much smaller than 1% (0.01), so reject H_0.
	Conclusion: there is strong evidence that birth months are not uniform: there is a bias for NHL players to be born earlier in the year (many more births than expected in Jan to Mar and fewer than expected in Oct to Dec).
5 (b)	H_0: the data can be modelled by the uniform distribution
	H_1: the data cannot be modelled by the uniform distribution

Question number	
	Expected frequency for each month = 75 ÷ 12 = 6.25 (all ≥ 5, so no classes combined)
	$X^2 = \Sigma(O - E)^2/E = \Sigma O^2/E - N$
	$\Sigma O^2 = 3^2 + 7^2 + 11^2 + 4^2 + 12^2 + 2^2 + 6^2 + 6^2 + 5^2 + 8^2 + 5^2 + 6^2 = 565$
	$X^2 = 565 \div 6.25 - 75 = 90.4 - 75 = 15.4$
	Degrees of freedom = 12 − 1 = 11
	10% critical value of $\chi^2_{11} = 17.275$
	Since 15.4 < 17.275, do not reject H_0.
	There is insufficient evidence to reject H_0: we can conclude that the birth months of Premier League footballers can be modelled by the uniform distribution.
6 (a)	The data suggest a non-linear relationship: the scatter diagram has three distinct sections, so the line underestimates income for some sizes (e.g. 2500 to 4000 sq ft) and overestimates for others (e.g. 1500 to 2500 sq ft).
6 (b)	$b = S_{xy} \div S_{xx} = 348\ 512\ 820.6 \div 2\ 869\ 673.03$
	$b = 121.4468746...$
	$a = \bar{y} - b\,\bar{x} = 3\ 907\ 142/37 - 121.4468746... \times 93\ 160/37$
	$a = 105\ 598.43... - 121.4468746... \times 2517.837...$
	$a = -200\ 185.1037...$
	$y = -200\ 185 + 121.4x$
7 (a)	A = (row total × column total) ÷ grand total = (115 × 115) ÷ 379
	A = 34.8945
	$B = (51 - 57.9551)^2 \div 57.9551 = 0.83467$
	$C = (6 - 16.6887)^2 \div 16.6887 = 6.84585$

Question number	
7 (b)	H_0: site of injury and sport are independent (no association) H_1: site of injury and sport are not independent (there is an association) Degrees of freedom = (7 − 1) × (3 − 1) = 12 5% critical value of χ^2_{12} = 21.026 Since 116.16 > 21.026, reject H_0. There is sufficient evidence at the 5% level to suggest that the site of injury and the sport are not independent.
7 (c)	Hand/Fingers, because of the very large contributions for Football (23.04) and Basketball (26.39). This is not surprising: there are disproportionately fewer hand/finger injuries in football (players rarely use their hands) and disproportionately more in basketball (the ball is handled all the time).
7 (d)	Misconception: the test also depends on the degrees of freedom, which are different here (df = (4 − 1)(3 − 1) = 6 instead of 12), so the two X^2 totals cannot be compared directly. She should compare the p-values of the two tests instead.