ADVANCED GCE

MATHEMATICS (MEI)   **4767**

Statistics 2

QUESTION PAPER

Candidates answer on the printed answer book.

OCR supplied materials:
* Printed answer book 4767
* MEI Examination Formulae and Tables (MF2)

Other materials required:
* Scientific or graphical calculator

Wednesday 22 June 2011
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the printed answer book and the question paper.

* The question paper will be found in the centre of the printed answer book.
* Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
* **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
* Use black ink. Pencil may be used for graphs and diagrams only.
* Read each question carefully. Make sure you know what you have to do before starting your answer.
* Answer **all** the questions.
* Do **not** write in the bar codes.
* You are permitted to use a scientific or graphical calculator in this paper.
* Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the printed answer book and the question paper.

* The number of marks is given in brackets **[ ]** at the end of each question or part question on the question paper.
* You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
* The total number of marks for this paper is **72**.
* The printed answer book consists of **12** pages. The question paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

* Do not send this question paper for marking; it should be retained in the centre or destroyed.

OCR is an exempt Charity

Turn over

1   An experiment is performed to determine the response of maize to nitrogen fertilizer. Data for the amount of nitrogen fertilizer applied, x kg/hectare, and the average yield of maize, y tonnes/hectare, in 5 experimental plots are given in the table below.

x	0	30	60	90	120
y	0.5	2.5	4.7	6.2	7.4

(i) Draw a scatter diagram to illustrate these data.   **[3]**

(ii) Calculate the equation of the regression line of y on x.   **[5]**

(iii) Draw your regression line on your scatter diagram and comment briefly on its fit. **[3]**

(iv) Calculate the value of the residual for the data point where $x = 30$ and $y = 2.5$. **[3]**

(v) Use the equation of the regression line to calculate estimates of average yield with nitrogen fertilizer applications of

(*A*) 45 kg/hectare,

(*B*) 150 kg/hectare.   **[2]**

(vi) In a plot where 150 kg/hectare of nitrogen fertilizer is applied, the average yield of maize is 8.7 tonnes/hectare. Comment on this result.   **[2]**

2   At a drive-through fast food takeaway, cars arrive independently, randomly and at a uniform average rate. The numbers of cars arriving per minute may be modelled by a Poisson distribution with mean 0.62.

(i) Briefly explain the meaning of each of the three terms 'independently', 'randomly' and 'at a uniform average rate', in the context of cars arriving at a fast food takeaway.   **[3]**

(ii) Find the probability of at most 1 car arriving in a period of 1 minute.   **[3]**

(iii) Find the probability of more than 5 cars arriving in a period of 10 minutes.   **[3]**

(iv) State the exact distribution of the number of cars arriving in a period of 1 hour. **[2]**

(v) Use a suitable approximating distribution to find the probability that at least 40 cars arrive in a period of 1 hour.   **[5]**

3 The weights of Braeburn apples on display in a supermarket, measured in grams, are Normally distributed with mean 210.5 and standard deviation 15.2.

    (i) Find the probability that a randomly selected apple weighs at least 220 grams. **[3]**

    (ii) These apples are sold in packs of 3. You may assume that the weights of apples in each pack are independent. Find the probability that all 3 of the apples in a randomly selected pack weigh at least 220 grams. **[2]**

    (iii) 100 packs are selected at random.

        (*A*) State the exact distribution of the number of these 100 packs in which all 3 apples weigh at least 220 grams. **[2]**

        (*B*) Use a suitable approximating distribution to find the probability that in at most one of these packs all 3 apples weigh at least 220 grams. **[5]**

        (*C*) Explain why this approximating distribution is suitable. **[1]**

    (iv) The supermarket also sells Cox's Orange Pippin apples. The weights of these apples, measured in grams, are Normally distributed with mean 185 and standard deviation σ.

        (*A*) Given that 10% of randomly selected Cox's Orange Pippin apples weigh less than 170 grams, calculate the value of σ. **[3]**

        (*B*) Sketch the distributions of the weights of both types of apple on a single diagram. **[4]**

[Question 4 is printed overleaf.]

4 **(a)** In a survey on internet usage, a random sample of 200 people is selected. The people are asked how much they have spent on internet shopping during the last three months. The results, classified by amount spent and sex, are shown in the table.

		Sex		Row totals
		Male	Female	
Amount spent	Nothing	28	34	62
	Less than £50	17	21	38
	£50 up to £200	22	26	48
	£200 up to £1000	23	16	39
	£1000 or more	8	5	13
Column totals		98	102	200

(i) Write down null and alternative hypotheses for a test to examine whether there is any association between amount spent and sex of person. **[1]**

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

		Sex	
		Male	Female
Amount spent	Nothing	0.1865	0.1791
	Less than £50	0.1409	0.1354
	£50 up to £200	0.0982	0.0944
	£200 up to £1000	0.7918	0.7608
	£1000 or more	0.4171	0.4007

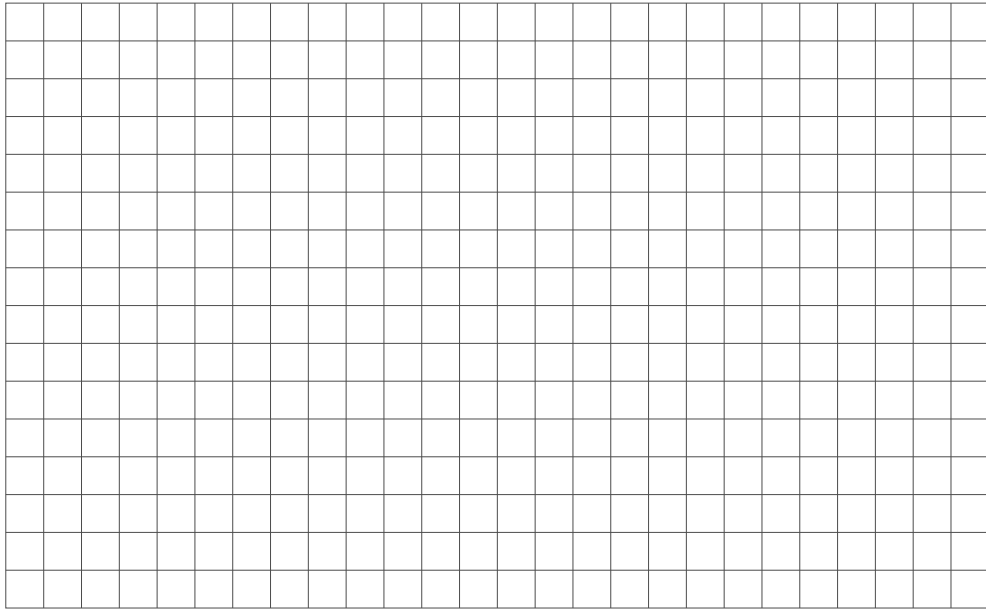The sum of these contributions, correct to 3 decimal places, is 3.205.

(ii) Calculate the expected frequency for females spending nothing. Verify the corresponding contribution, 0.1791, to the test statistic. **[3]**

(iii) Carry out the test at the 5% level of significance, stating your conclusion clearly. **[5]**

(b) A bakery sells loaves specified as having a mean weight of 400 grams. It is known that the weights of these loaves are Normally distributed and that the standard deviation is 5.7 grams. An inspector suspects that the true mean weight may be less than 400 grams. In order to test this, the inspector takes a random sample of 6 loaves. Carry out a suitable test at the 5% level, given that the weights, in grams, of the 6 loaves are as follows. **[9]**
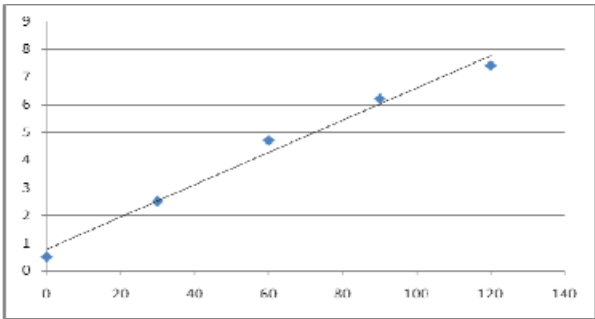
$$392.1 \qquad 405.8 \qquad 401.3 \qquad 387.4 \qquad 391.8 \qquad 400.6$$
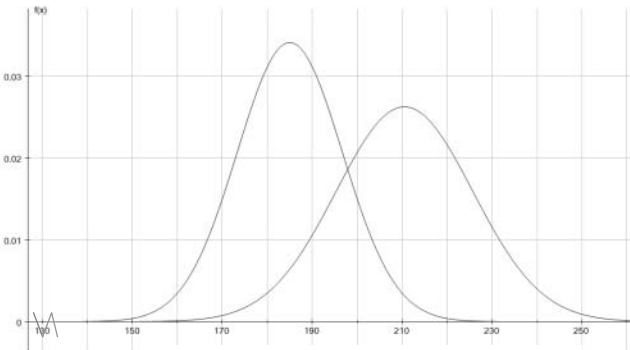
1 (i)

1 (ii)

1 (i)		G1 for axes G1 For values of x G1 for values of y		Condone axes drawn either way. Axes should show some indication of scale. If not then Max G1 if points 'visibly correct'. If axes are scaled and only one point is incorrectly plotted, allow max G2/3.
			3	
1 (ii)	$\bar{x} = 60,\ \bar{y} = 4.26$ $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{1803 - 300 \times 21.3/5}{27000 - 300^2/5} = \dfrac{525}{9000} = 0.0583$ OR $\ b = \dfrac{1803/5 - 60 \times 4.26}{27000/5 - 60^2} = \dfrac{105}{1800} = 0.0583$ hence least squares regression line is: $\qquad y - \bar{y} = b(x - \bar{x})$ $\Rightarrow \quad y - 4.26 = 0.0583\,(x - 60)$ $\Rightarrow \qquad y = 0.0583x + 0.76$	B1 for $\bar{x}$ and $\bar{y}$ used appropriately (SOI) M1 for attempt at gradient (b) A1 for 0.0583 **cao** M1 for equation of line A1 FT for complete equation		B1 for means can be implied by a correct value of b using either method. Allow $\bar{y} = 4.3$ Attempt should be correct – e.g. evidence of either of the two suggested methods should be seen. Allow 0.058 Condone $0.058\dot{3}$ and $\dfrac{7}{120}$ Dependent on first M1. Values must be substituted to earn M1. Condone use of their b for FT provided $b>0$. Final equation must be simplified. $b = 0.058$ leads to $y = 0.058x + 0.78$
			5	
1 (iii)	Regression line plotted on graph The fit is good	G1 G1 E1 for good fit		Line must pass through their $(\bar{x}, \bar{y})$ and y-intercept. E0 for notably inaccurate graphs/lines
			3	

1 (iv)	$x = 30 \Rightarrow$ predicted $y = 0.0583 \times 30 + 0.76 = 2.509$ Residual $= 2.5 - 2.509 = -0.009$	B1 for prediction M1 for subtraction A1 FT	**3**	Using their **equation** Subtraction can be 'either way' but for the final mark the sign of the residual must be correct. FT sensible equations only – e.g. no FT for $y = 0.071x$ leading to $+0.37$. [$c = 0.78$ leads to a residual of -0.02]

1 (v)	(*A*) For $x = 45$,
	       $y = 0.0583 \times 45 + 0.76 = 3.4$
	
	 (*B*) For $x = 150$,
	       $y = 0.0583 \times 150 + 0.76 = 9.5$
		M1 for at least one prediction attempted
		
		 A1 for **both answers** (FT their equation provided their $b>0$)
			2	Prediction obtained from their equation.

1 (vi)	This is **well** below the predicted value … …suggesting that the model breaks down for larger values of *x*.	E1 for **well** below E1 extrapolation	**2**	Some indication that the value (8.7) is **significantly** below what is expected (9.5) is required for the first E1. Simply pointing out that it is 'below' is not sufficient. The second E1 is available for a suitable comment relating to the model being suitable only for values within the domain of the given points. Allow other sensible comments for either E1. E.g. The data might be better modelled by a curve', 'there may be other factors affecting yield',
			18	

2 (i)	Independently means that the arrival time of each car is unrelated to the arrival time of any other car.	E1		NOTE **Each answer must be 'in context' and 'clear'**
	Randomly means that the arrival times of cars are not predictable.	E1		Allow sensible alternative wording.
	At a uniform average rate means that the average rate of car arrivals does not vary over time.	E1	**3**	SC1 For ALL answers not in context but otherwise correct.
2 (ii)	P(At most 1 car) $= e^{-0.62}\dfrac{0.62^0}{0!} + e^{-0.62}\dfrac{0.62^1}{1!}$	M1 for either		$1.62e^{-0.62}$
		M1 for sum of both		Allow 0.8715 not 0.872 or 0.8714
	$= 0.5379\ldots + 0.3335\ldots = 0.871$	A1 CAO	**3**	Allow 0.87 without wrong working seen
2 (iii)	New $\lambda = 10 \times 0.62 = 6.2$	B1 for mean (SOI)		
		M1 for probability		Use of $1 - P(X \le 5)$ with any λ
	P(more than 5 in 10 mins) $= 1 - 0.4141 = 0.5859$	A1 CAO	**3**	Allow 0.586
2 (iv)	Poisson with mean 37.2	B1 for Poisson		Dependent on first B1
		B1 for mean 37.2	**2**	Condone P(37.2, 37.2)
2 (v)	Use Normal approx with $\mu = \sigma^2 = \lambda = 37.2$	B1 for Normal (SOI)		
		B1 for parameters		
	$P(X \ge 40) = P\left(Z > \dfrac{39.5 - 37.2}{\sqrt{37.2}}\right)$	B1 for 39.5		
	$= P(Z > 0.377) = 1 - \Phi(0.377) = 1 - 0.6469$	M1 for correct use of Normal approximation using correct tail		
	$= 0.3531$	A1 **cao**	**5**	Allow 0.353
			16	

3 (i)	P(Apple weighs at least 220g) $= P\left(Z > \dfrac{220 - 210.5}{15.2}\right)$ $= P(Z > 0.625)$ $= 1 - \Phi(0.625) = 1 - 0.7340$ $= 0.2660$	M1 for standardising M1 for correct structure A1 CAO inc use of diff tables	**3**	Condone numerator reversed but penalise continuity corrections i.e. $1 - \Phi$(positive z value) Allow 0.266 but not 0.27
3 (ii)	P(All 3 weigh at least 220g) $= 0.2660^3 = 0.0188$	M1 A1 FT	**2**	M1 for their answer to part (i) cubed Allow 0.019 and 0.01882
3 (iii)	(*A*) Binomial (100, 0.0188) (*B*) Use a Poisson distribution with $\lambda = 100 \times 0.0188 = 1.88$ $P(\text{At most one}) = e^{-1.88}\dfrac{1.88^0}{0!} + e^{-1.88}\dfrac{1.88^1}{1!}$ $= 0.1525 + 0.2869 = 0.4394$ (*C*) n is large and p is small	B1 for binomial B1 for parameters B1 for Poisson SOI B1 for Poisson mean M1 for either probability M1 for sum of both A1 CAO For 0.44 or better B1	**2** **5** **1**	Second B1 dependent on first B1 FT their answer to part (ii) for second B1 Consistent with $p < 0.1$ from part (iii) (*A*) FT answer to part (ii) with $p < 0.1$ Dependent on both previous B1 marks Allow 0.4395 but not 0.4337 Dependent on use of Poisson in part (iii) *B* Allow n is large and $np <10$ & n is large and $np \approx npq$
3(iv)(*A*)	$\Phi^{-1}(0.1) = -1.282$ $\dfrac{170 - 185}{\sigma} = -1.282$ $1.282\,\sigma = 15$ $\sigma = 11.70$	B1 for ±1.282 M1 for correct equation as written o.e. A1 CAO	**3**	Do not allow $1 - 1.282$ Allow M1 if different *z*-**value** used Without incorrect working seen. Allow 11.7

3(iv)(*B*)	Cox's Braeburns	G1 for shape G1 for means, shown explicitly or by scale G1 for lower max height for Braeburns G1 for greater width (variance) for Braeburns	**4**	Ignore labelling of vertical axis. Two intersecting, adjacent Normal curves Means at 185 and 210.5
		TOTAL	**20**	
4(a)(i)	H_0: no association between amount spent and sex H_1: some association between amount spent and sex	B1 for both	**1**	Hypotheses must be the right way round, in context and must not mention 'correlation'.
4(a)(ii)	Expected frequency = $62 \times 102 \div 200 = 31.62$ Contribution = $(34 - 31.62)^2 / 31.62$ = 0.1791	B1 M1 A1 for valid attempt at $(O-E)^2/E$ *NB Answer given*	**3**	Do not allow 31.6

4(a)(iii)	Refer to X_4^2 Critical value at 5% level = 9.488 3.205 < 9.488 Result is not significant There is insufficient evidence to suggest any association between amount spent and sex. .	B1 for 4 deg of freedom B1 CAO for cv M1 A1 for not significant E1	**5**	Allow $p = 0.524$ 0.524 > 0.05 Conclusion must be stated to earn A1 here. Allow 'do not reject H_0' and condone 'accept H_0' or 'reject H_1'. FT if cv consistent with their d.o.f. Dependent on previous A1 and final comment must be in context and not mention correlation. SC1 for correct final conclusion where previous A1 omitted but M1 awarded.
4 (b)	H_0: $\mu = 400$; H_1: $\mu < 400$ Where μ denotes the population mean (weight of the loaves). $\bar{x} = 396.5$ Test statistic $= \dfrac{396.5 - 400}{5.7/\sqrt{6}} = \dfrac{-3.5}{2.327} = -1.504$ 5% level 1 tailed critical value of z = −1.645 −1.504 > −1.645 so not significant. There is insufficient evidence to reject H_0 There is insufficient evidence to suggest that the true mean weight of the loaves is lower than the minimum specified value of 400 grams.	B1 for H_0 B1 for H_1 B1 for definition of μ B1 for sample mean M1 must include $\sqrt{6}$ A1FT their sample mean B1 for ±1.645 M1 for sensible comparison leading to a conclusion A1 for conclusion in context	**9**	Hypotheses in words must refer to population mean. Condone numerator reversed for M1 but award A1 only if test statistic of 1.504 is compared with a positive z-value. Dependent on previous M1 FT their sample mean only if hypotheses are correct.
		TOTAL	**18**	

Additional notes re Q1 parts (ii), (iv) and (v)

Part (ii) 'x on y' max B1
Part (iv) $x = 16.9y – 12.02$ leads to a prediction of $x = 30.23$ and a residual of -0.23 B1 M1 A1 available.
Part (v) 'x on y' not appropriate here so award 0 if 'x on y' used.

Additional notes re Q2 parts (i) & (v)
Part (i)
Independent – Allow 'not linked to' or 'no association' or 'unrelated' 'not affected by', 'not connected to', 'not influenced by'
DO NOT ACCEPT 'not together' or 'not dependent'
Random – Allow 'not predictable' or 'no pattern' or 'could happen at any time' or 'not specific time'
Uniform average rate – Allow 'average (rate) is constant over time' DO NOT ACCEPT 'average constant' or 'average rate and uniform' – be generous over defining 'average' and/or 'rate'.
Part (v) If Binomial distribution stated in part (iv), allow B1 B0 B1 M0 A0 max

Additional notes re Q3 part (iii) where $p > 0.1$
(iii) B – as scheme unless a Normal approximation is more suitable ($p > 0.1$). If so, award B1 B1 for Normal and correct parameters. The remaining marks are dependent on both these B1 marks being awarded. M1 for the correct continuity correction (P($X < 1.5$)) and depM1 for the correct tail but award A0.
(iii) C – 'n is large and p is not too small' or '$np > 10$'

Additional notes re Q4(b)

σ estimated
sample mean, 7.079… used in place of 5.7, the given value of the population mean, leads to a test statistic of -1.212… This gets M1A0 & the remaining marks are still available.

Critical Value Method
$400 - 1.645 \times 5.7 \div \sqrt{6}$… gets M1B1 …= 396.17…gets A1
$400 + 1.645 \times 5.7 \div \sqrt{6}$ gets M1B1A0.
396.5 > 396.2 gets M1 for sensible comparison (and B1 for 396.5)
A1 still available for correct conclusion in words & context

90% Confidence Interval Method
CI centred on 396.5 (gets B1 for 396.5)
+ or $- 1.645 \times 5.7 \div \sqrt{6}$ gets M1 B1
= (392.67, 400.33) A1
contains 400 gets M1
A1 still available for correct conclusion in words & context

Probability Method
Finding P(sample mean < 396.5) = 0.0663 gets M1 A1 (and B1 for 396.5)
0.0663 > 0.05 gets M1 for a sensible comparison if a conclusion is made and also gets the B1 for 0.0663 (to replace the B1 for cv = 1.645).
A1 still available for correct conclusion in words & context.
Condone P(sample mean >396.5) = 0.9337 for M1 and B1 for 0.9337 but only allow A1 if later compared with 0.95 at which point the final M1and A1 are still available

Two-tailed test
Max B1 B0 B1 B1 M1 A1 B1 (for cv = -1.96) M1 A0