edexcel

Mark Scheme (Results)

Summer 2016

Pearson Edexcel GCE Statistics S3 (6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2016
Publications Code 6691_01_1606_MS

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75

2. The Edexcel Mathematics mark schemes use the following types of marks:

   - **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
   - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
   - **B** marks are unconditional accuracy marks (independent of M marks)
   - Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol $\surd$ will be used for correct ft
   - cao – correct answer only
   - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
   - isw – ignore subsequent working
   - awrt – answers which round to
   - SC: special case
   - oe – or equivalent (and appropriate)
   - d… or dep – dependent
   - indep – independent
   - dp decimal places
   - sf significant figures
   - $*$ The answer is printed on the paper or ag- answer given
   - $\lfloor$ or d… The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
   - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
   - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

June 2016
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
1(a) e.g.	Analyse / find estimates for a particular **subgroup** of the population. Stratified guarantees **representation** of all **groups**, srs does not. Observe **relationships between subgroups** – srs does not guarantee equal or proportionate representation. Rare or extreme cases as part of a **small subgroups** can be represented **proportionately** in stratified i.e. stratified represents the structure of the population– srs does not allow this. Stratified typically require **large sample size** compared to srs due to lower variability within subgroups compared to entire population. Any 2 distinct reasons	B1B1 (2)
(b) e.g.	It (a stratified sample) is **not biased** as the members are chosen randomly. You can **estimate the sampling errors** (for a stratified sample) It (a stratrified sample) gives **more accurate estimates** as it is a random process. A quota sample may be (**interviewer / process**) **biased**. It's **not possible to estimate/find the sampling errors** for a quota sample (whereas you can for a stratified sample) Any 2 distinct reasons	B1B1 (2) **Total 4**
Notes	Award B1B1 two correct, B1B0 one correct. Allow 'it' for 'stratified'. Do not award marks for vague responses such as 'cheap', 'easy' 'quick' 'random' etc. Mentioning 'sampling frame' alone is not sufficient for a mark. Mentioning 'non-response are not recorded' alone is not sufficient for a mark.	

Question Number	Scheme	Marks
2	H_0 : **Drug concentration** and catching **influenza** are independent / not associated	
	H_1 : **Drug concentration** and catching **influenza** are not independent / associated	B1

	A	*B*	*C*	
Influenza	$\frac{50\times 27}{110}=12.272...$	$\frac{50\times 52}{110}=23.636...$	$\frac{50\times 31}{110}=14.090...$	50
No influenza	$\frac{60\times 27}{110}=14.727...$	$\frac{60\times 52}{110}=28.363...$	$\frac{60\times 31}{110}=16.909...$	60
	27	52	31	110

M1A1

O	*E*	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
12	12.272…	0.0060…	11.7333…
29	23.636…	1.2171…	35.5807…
9	14.090…	1.8392…	5.7483…
15	14.727…	0.0050…	15.2777…
23	28.363…	1.0142…	18.6506…
22	16.909…	1.5327…	28.6236…

M1A1

$\sum\frac{(O-E)^2}{E}=5.6145...$ or $\sum\frac{O^2}{E}-N=115.62...-110=5.6145...$ awrt 5.61-5.62 A1

$\nu=(3-1)(2-1)=2$, $\chi^2_2(10\%)=4.605$ B1B1ft

Reject H_0 M1

Drug concentration and catching **influenza** are not independent / are associated. A1cao

(10)

Total 10

Notes

B1 hyps correct way around

M1 for correct expression at least once

A1 **all seen** and correct 2dp or better. Can be implied by test statistic of awrt 5.61-5.62.

M1 either method at least one correct

A1 at least 3 correct values. Can be implied by test statistic of awrt 5.61-5.62

A1 awrt 5.61-5.62

B1 cao

B1 follow through their ν

M1 must correctly reject / not reject the null hypothesis based on their test stat and cv oe

Question Number	Scheme	Marks
3 (a)	Variables do not have a (joint) normal distribution Relationship is not linear The given data is ordinal Any 1	B1 (1)
(b)	<table><tr><td>Day</td><td>Sun</td><td>Mon</td><td>Tues</td><td>Weds</td><td>Thurs</td><td>Fri</td><td>Sat</td></tr><tr><td>Ice-cream</td><td>6</td><td>4</td><td>7</td><td>5</td><td>3</td><td>2</td><td>1</td></tr><tr><td>Sunglasses rank</td><td>6</td><td>5</td><td>7</td><td>2</td><td>3</td><td>4</td><td>1</td></tr><tr><td>d</td><td>0</td><td>-1</td><td>0</td><td>3</td><td>0</td><td>-2</td><td>0</td></tr><tr><td>d^2</td><td>0</td><td>1</td><td>0</td><td>9</td><td>0</td><td>4</td><td>0</td></tr></table>	
	$\sum d^2 = 14$	M1
	$r_s = 1 - \frac{6\times 14}{7(49-1)} = 0.75$	M1A1 (3)
(c)	$H_0 : \rho = 0$, $H_1 : \rho > 0$	B1
	5% cv 0.7143	B1
	Reject H_0	M1
	Evidence of **positive** correlation between sales of **ice cream** and sales of **sunglasses.**	A1cao (4)
(d)	(5% cv 0.6694)	
	Accept H_0	M1
	Insufficient evidence of **positive** correlation between sales of ice cream and sales of sunglasses.	A1cao (2)
(e)	Suggests relationship might be non-linear.	B1 (1) **Total 11**

Notes

(a) Accept ‘already ranked’ oe for ordinal
Accept one variable is not normally distributed

(b) M1 attempt to find d, d^2 and sum. may be implied by sight of $\sum d^2 = 14$
M1 for use of the correct formula, follow through their $\sum d^2$ if clearly stated. If answer is not correct, a correct expression is required.
A1 0.75 cao

(c) 1st B1for both hypotheses in terms of ρ, one tail. Allow use of ρ_s.
Only award if no errors seen in hypotheses in part(c) **and part(d)**
Hypotheses just in words e.g. “no correlation” score B0.
B1 0.7143 cao
M1 must correctly reject / not reject the null hypothesis based on their test stat and cv oe
A1 Conclusion must mention ice cream and sunglasses

(d) M1 for not rejecting / accepting null hyp
A1 must mention ice cream and sunglasses

Question Number	Scheme	Marks
4 (a)	X_i be rv 'weight of i^{th} randomly chosen egg'	
	$\mathrm{E}(X_1 - X_2) = 0$	B1
	$\mathrm{Var}(X_1 - X_2) = 2\times 5^2 = 50$	B1
	$\mathrm{P}(\lvert X_1 - X_2\rvert > 2) = 2\mathrm{P}(X_1 - X_2 > 2)$	M1
	$= 2\mathrm{P}(Z > \frac{2}{\sqrt{50}})$	dM1
	$= 2\mathrm{P}(Z > 0.2828...)$	
	$= 2(1-0.6103)=0.7794$ awrt 0.777-0.779	A1
		(5)
(b)	$W = C + X_1 + X_2 + ... + X_{12}$	
	$\mathrm{E}(W) = 40 + 12\times 60 = 760$	B1
	$\mathrm{Var}(W) = 1.5^2 + 12\times 5^2$	M1
	$= 302.25$	A1
	Distribution is $\mathrm{N}(760, 302.25)$	
		(3)
(c)	$\mathrm{P}(W > 800) = \mathrm{P}\left(Z > \frac{800-760}{\sqrt{302.25}}\right)$	M1
	$= 1 - \mathrm{P}(Z < 2.3007...)$	
	$= 0.0107$ awrt 0.0107	A1
		(2)
		Total 10
Notes (a)	B1 for 0	
	B1 for 50	
	M1 for $\lvert X_1 - X_2\rvert > 2$ seen. Accept $X_1 - X_2 > 2$ provided a subsequent doubling of the probability is seen. i.e. 0.3897 x 2.	
	dM1 standardise with their 0 and their $\sqrt{50}$ dependent on previous M.	
	A1 awrt 0.777-0.779	
(b)	B1 for 760	
	M1 requires squares	
	A1 cao	
(c)	Must be finding correct probability (ie $\mathrm{P}(W > 800)$ or $\mathrm{P}(Z > 2.3007...)$ etc) and standardise with 800 and their 760 and their $\sqrt{302.25}$	
	A1 awrt 0.0107 from correct working.	

Question Number	Scheme	Marks
5(a)	$H_0 : \mu_e = \mu_n$, $H_1 : \mu_e > \mu_n$	B1
	$z = \dfrac{26.3-24.8}{\sqrt{\dfrac{12.2}{35}+\dfrac{10.1}{42}}} = \dfrac{1.5}{\sqrt{0.58904...}} = \dfrac{1.5}{0.76749...}$	M1M1
	$z = 1.9544...$ awrt 1.95	A1
	Critical value is 1.6449	B1
	Reject H_0. Doctor's claim is supported.	A1
		(6)
(b)	Either assume $\bar{X}$ has a normal distribution (for both samples) or assume sample sizes are large enough to use CLT Assume individual **results** are independent	
	Assume $\sigma^2 = s^2$ for **both** populations or a single general population	B1 B1
		(2)
(c)	$\bar{x} = \left(\dfrac{35\times 26.3+31.7}{36} = \dfrac{952.2}{36}\right) 26.45$	B1
	For $n = 35$, $\sum x^2 = 34\times 12.2 + 35\times 26.3^2 (= 24623.95)$	M1
	For $n = 36$, $s^2 = \dfrac{25628.84 - 36\times 26.45^2}{35} = 12.661...$ awrt 12.7	dM1A1
		(4)
		Total 12
Notes (a)	Both hyps, one tailed only oe.	
	Accept μ_1, μ_2 or μ_A, μ_B etc if there is some indication of which is which.	
	M1 for correct method for standard error	
	M1 for whole expression	
	A1 awrt 1.95	
	B1 1.6449 or $p = 0.974...$ (>0.95)	
	A1 must mention doctor and claim or description of claim that includes '**mean** lung capacity' and 'exercise'.	
ALT (a)	M1 for $\sqrt{\dfrac{12.2}{35}+\dfrac{10.1}{42}}$	
	M1 for $1.6449 = \dfrac{c}{\sqrt{\dfrac{12.2}{35}+\dfrac{10.1}{42}}}$	
	A1 for awrt $c = 1.26$ seen	
	B1 1.5	
(c)	M1 Attempt $\sum x^2 = 34\times 12.2 + 35\times 26.3^2$	
	or $\sum (x-\bar{x})^2 = 34\times 12.2 + 35(26.45-26.3)^2 (= 415.5875)$	
	dM1 $s^2 = \dfrac{\sum x^2 + 31.7^2 - 36\times 26.45^2}{35}$ or $s^2 = \dfrac{415.5875 + (31.7-26.45)^2}{35}$	
	A1 awrt 12.7	

Question Number	Scheme	Marks
6(a)	H_0 : Binomial with $p = 0.3$ is a good fit. H_1 : Binomial with $p = 0.3$ is not a good fit.	B1
		M1A1
	$\sum \frac{(O-E)^2}{E} = 4.097...$ or $\sum \frac{O^2}{E} - N = 54.097... - 50 = 4.097...$ awrt 4.09-4.1(0)	dM1A1
	$\nu = 3-1 = 2$	B1ft
	$\chi^2_2(5\%) = 5.991$ $(> 4.1(0))$	B1ft
	Insufficient evidence to reject H_0 (Accept H_0)	
	Binomial with $p = 0.3$ is a good fit.	A1
		(8)
(b)	$\overline{x} = \frac{40+62+54+24}{100} = 1.8$	B1 cao
	$r = 26.78$	B1 cao
	$s = 16.07$	B1 cao
		(3)
(c)	H_0 : Poisson is a good fit. H_1 : Poisson is not a good fit.	B1
	$\sum \frac{(O-E)^2}{E} = 14.65 - 14.66$ or $\sum \frac{O^2}{E} - N = 114.65 - 100 = 14.65 - 14.66$	M1A1
	$\nu = 5-1-1 = 3$	B1 cao
	$\chi^2_3(1\%) = 11.345$ (< 14.65)	B1ft
	Sufficient evidence to reject H_0	
	Poisson is not a good fit.	A1 cao
		(6)
		Total 17

Notes (a)	B1 both including $p = 0.3$
	M1 with some combined columns and at least one *E* correct to 2sf
	A1 all correct to 2dp and total of expected values is 50.
	dM1 either method
	A1 awrt 4.09-4.1(0)
	B1 ft their columns -1
	B1 ft their
	A1 cao
(c)	B1 no parameters included
	M1 either method

	0	1	2 or more
Observed	6	25	19
Expected	50x0.2401 =12.005 or **12.01** or 12.00	50x0.4116 =20.58	50x0.2646+50x0.0756+50x0.0081 =13.23+3.78+0.405 =17.415 or **17.41** or 17.42
$\frac{(O-E)^2}{E}$	3.003751	0.949291	0.144256
$\frac{O^2}{E}$	2.998751	30.36929	20.72926

	0	1	2	3	4 or more
Observed	5	40	31	18	6
Expected	16.53	29.75	26.78	16.07	10.87
$\frac{(O-E)^2}{E}$	$\frac{11.53^2}{16.53} = 8.042...$	$\frac{10.25^2}{29.75} = 3.532...$	$\frac{4.22^2}{26.78} = 0.665...$	$\frac{1.93^2}{16.07} = 0.232...$	$\frac{4.87^2}{10.87} = 2.182...$
$\frac{O^2}{E}$	$\frac{5^2}{16.53} = 1.512...$	$\frac{40^2}{29.75} = 53.782...$	$\frac{31^2}{26.78} = 35.885...$	$\frac{18^2}{16.07} = 20.162...$	$\frac{6^2}{10.87} = 3.312...$

	B1 ft their ν	
Question Number	Scheme	Marks
7(a)	$19.5 \pm 1.6449 \times \frac{1.5}{\sqrt{50}}$	M1B1
	$= (19.151..., 19.848...)$ awrt 19.2, awrt 19.8	A1A1
		(4)
(b)	CI does not contain 20 oe	M1
	Fast Food restaurant statement is too high; they should reduce the stated value.	A1
		(2)
(c)	$P(\lvert \bar{X} - \mu \rvert < 0.5) = 0.9$	
	$\frac{0.5}{\frac{2}{\sqrt{n}}} = 1.6449$	M1A1
	$n = \left(2 \times \frac{1.6449}{0.5} \right)^2 = 43.29...$	dM1A1
	Sample size required is 44	A1
		(5)
		Total 11
Notes		
(a)	M1 correct with their z i.e. $19.5 \pm (z \text{ value}) \times \frac{1.5}{\sqrt{50}}$	
	B1 for 1.6449	
	A1 awrt 19.2, A1 awrt 19.8(5)	
(b)	M1 Require 20 compared to their interval	
	A1 Accept statement that relates to 20 being above the interval.	
(c)	M1 $\frac{0.5}{\frac{2}{\sqrt{n}}} = z$ value or equivalent expression	
	A1 All correct	
	dM1 Attempt to solve $\frac{0.5}{\frac{2}{\sqrt{n}}} =$ their z value	
	A1 awrt 43.3	
	A1 44 cao	