edexcel

Mark Scheme (Results)

Summer 2016

Pearson Edexcel International A Level Statistics 3

(WST03/01)

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 WST03_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.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL IAL 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
   - $\sqsubset$ 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 IAL
WST03/01 Statistics 3
Mark Scheme

Question Number	Scheme	Marks
1. (a)		

Salesperson	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
Rank Distance	7	6	4	1	5	3	2	8
Rank Commission	8	5	7	3	1	2	6	4

or

Salesperson	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
Rank Distance	2	3	5	8	4	6	7	1
Rank Commission	1	4	2	6	8	7	3	5

M1

$\sum d^2 = 1+1+9+4+16+1+16+16;\ = 64$ M1

$\sum d^2 = 64$ A1

$r_s = 1 - \frac{6(64)}{8(63)};\ = 0.238095...$ dM1;

$\frac{5}{21}$ or awrt **0.238** A1

[5]

(b) $H_0 : \rho = 0$, $H_1 : \rho > 0$ B1

Critical Value $r_s = 0.6429$ or CR: $r_s \geqslant 0.6429$ Critical value of **0.6429** B1

Either
- Do not reject H_0 (accept H_0)
- Result is not significant
- $r_s = 0.238...$ does not lie in the CR

see notes M1

conclude that there is no positive correlation between distance travelled and amount of commission received. A1

[4]

9

Notes

(a)

1st M1 For an attempt to rank at least one row (at least four correct)

2nd M1 For an attempt at d^2 row for their ranks (may be implied by $\sum d^2 = 64$)

1st A1 $\sum d^2 = 64$ (May be implied by correct answer)

3rd dM1 dependent on 1st M1 for use of $1 - \frac{6\sum d^2}{8(63)}$ with their $\sum d^2$

(b)

1st B1 Both hypotheses stated in terms of ρ or ρ_s.

M1 For a correct non-contradictory statement relating their r_s with their c.v. where $|\text{their c.v.}| < 1$
e.g. 'Do not reject H_0', 'not significant', 'not in critical region'

A1 Dependent on all previous marks in (b) scored.
For a contextualised comment which is accepting H_0, which must mention "no positive correlation", "distance" and "commission". (Use of "association" only is A0.)
Follow through their r_s with 0.6429 (provided |their r_s| < 1)

Note **Two-tailed test**
Applying a two-tailed test scores a maximum of B0B1M1A0
So Award SC B0B1 for $H_0 : \rho = 0$, $H_1 : \rho \neq 0$ followed by critical value $r_s = (\pm)\ 0.7381$ and allow access to the M1 mark only.

Question Number	Scheme	Marks
2. (a)	H_0 : There is no association between centre and result (independent) H_1 : There is an association between centre and result (dependent)	Correct hypotheses B1
		Some attempt at $\frac{\text{(Row Total)(Column Total)}}{\text{(Grand Total)}}$ Can be implied by at least one correct E_i to 1d.p. M1
		All expected frequencies are correct to awrt/trunc. 2dp. A1
		At least 2 correct terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. Accept 2 sf accuracy for the dM1 mark. dM1
		At least 5 correct $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ terms to either 2 dp or better. Allow truncation. A1
	$X^2 = \sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 620$;= awrt 7.52	For applying either $\sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 620$ dM1
		$X^2 = 7.519087...$ awrt **7.52** A1
	$\nu = (2-1)(3-1) = 2$	$\nu = 2$ (can be implied) B1
	$\chi^2_2(0.05) = 5.991 \Rightarrow$ CR: $X^2 \geqslant 5.991$ [in the CR/significant/Reject H_0]	**5.991** B1
	conclude that there is an association between driving test centre and result. (or they are not independent.)	A1 **[10]**
(b)	Test centre *C*	B1
	Observed and expected differences are bigger for test centre *C* than for any other test centre/ Centre *C* contributes most to the test statistic/**Pass rate** at *C* is lower than the **pass rate** at the other centres.	dB1 **[2]**
		12

Observed	Expected	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
99	92.48	0.4596...	105.9796...
110	100.97	0.8075...	119.8375...
68	83.55	2.8941...	55.3441...
108	114.52	0.3712...	101.8512...
116	125.03	0.6521...	107.6221...
119	103.45	2.3373...	136.8873...
	Totals	7.522	627.522...

Expd	*A*	*B*	*C*	Total
Pass	92.482...	100.970...	83.546...	(277)
Fail	114.517...	125.029...	103.453...	(343)
Total	(207)	(226)	(187)	(620)

Notes

(a)

1st B1 For both hypotheses. Must mention "centre" **and** "result" at least once.
Use of "relationship" or "correlation" or "connection" is B0.

2nd dM1 Dependent on 1st M1 for at least 2 correct terms or correct expressions with their E_i

2nd A1 At least 5 correct terms to either 2 d.p. or better. Allow truncated answers. (May be implied).

3rd dM1 Dependent on 2nd M1
For applying either $\sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 620$

Note Correct answer **without** the expected frequencies stated scores **special case** M0A0M1A1M1A1

2nd B1 $\nu = 2$ This mark can be implied by a correct critical value of 5.991

4th A1 Dependent on the 3rd M1 and 3rd B1. A correct contextualised conclusion which is rejecting H_0.
Must mention "centre" **and** "result"oe
If hypotheses are the wrong way round, then A0 here.
Contradictory statements score A0. E.g. "significant, do not reject H_0".
Condone "relationship" or "connection" here but **not** "correlation".

Question Number	Scheme	Marks
3. (a)	Any two reasons from • sample will be taken from the same office **or** other offices not considered. • same day **or** other days not considered. • around the same time of arrival **or** first 50 employees • These employees may have the same views (e.g. positive attitude to work).	B1, B1 **[2]**
(b)	Generate a numbered list(oe) of all employees sorted by office location.	B1
	Use random numbers to select/take a (simple) random sample of ...	B1
	51 employees from Bristol, 26 employees from Dudley, 73 employees from Glasgow.	B1cao **[3]**
(c)	Any one of advantage of stratified sampling, e.g. • A stratified sample is not biased as the members are chosen randomly. • You can estimate the sampling errors for a stratified sample • A stratified sample gives more accurate estimates as it is a random process.	B1 **[1]** **6**
	Notes	
(a)	B1B0 for one suitable reason B1B1 for two suitable reasons	
(b)	1st B1 for a suitable numbered/labelled list for each region 2nd B1 for use of random numbers/sample to select employees 3rd B1 for 51 with Bristol, 26 with Dudley and 73 with Glasgow	
(c)	Note Allow 'it' for stratified sample B0 for "a stratified sample can reflect the population structure." B0 for "estimates obtained from each of the strata."	

Question Number	Scheme	Marks
4. (a)	$H_0 : \mu_C = \mu_A \quad H_1 : \mu_C > \mu_A$	B1
	s.e. $= \sqrt{\frac{5.9^2}{60} + \frac{5.2^2}{50}}$ $\{= 1.058757133...\}$	M1
	$z = \frac{61.2 - 59.1}{'1.0587...'}; = 1.983457711$ awrt $\pm$ **1.98**	dM1; A1
	One tailed c.v. $Z = 1.6449$ **or** CR: $Z \geqslant 1.6449$ **or** p-value = awrt $0.024 < 0.05$ [in the CR/significant/Reject H_0 / "0.024" < 0.05]	B1
	Conclude that the mean time taken by children to complete a task is greater than that of adults.	A1
		[6]
(b)	$\bar{X}_C$ and $\bar{X}_A$ are both approximately normally distributed.	B1
		[1]
(c)	Have assumed $s^2 \simeq \sigma^2$ / variance of sample $\simeq$ variance of population	B1
		[1]
		8

Notes

(a)

1st B1 If μ_1, μ_2 used then it must be clear which refers to children/adults.

Note Also allow $H_0 : \mu_C - \mu_A = 0$ $H_1 : \mu_C - \mu_A > 0$

1st M1 s.e. $= \sqrt{\frac{5.9^2}{60} + \frac{5.2^2}{50}}$. (may be implied by s.e. = awrt 1.06)

Condone minor slips e.g. $\sqrt{\frac{5.9^2}{50} + \frac{5.2^2}{60}}$

2nd dM1 Dependent on 1st M1. (Allow $\pm$)
Follow through their s.e. if 1st M1 mark has been awarded.

2nd B1 For 1.6449 (compatible with sign of their test statistic) or correct probability comparison. (Condone: "0.976" > 0.95)

2nd A1 Dependent on both method marks being scored and for rejecting H_0
For a correct conclusion in context which is based on their *z*-value and their critical value, where $|\text{c.v.}| > 1$
Contradictory statements score final A0. E.g. "significant, do not reject H_0."

(a) **Alternative method for 2nd "M1A1B1" marks:** Let $D = \bar{x}_C - \bar{x}_A$

$1.6449 = \frac{D - 0}{1.0587...}$

So, $D = 1.741...$

dM1: dependent on the 1st M1 for $\frac{D}{\text{their "1.0587..."}} = 1.6449 / 1.645 / 1.64 / 1.65$

A1: D = awrt 1.74

B1: 1.6449

(b) Allow in words e.g "**sample means** are normally distributed"

(c) Allow $s = \sigma$ but watch out for $s_C = s_A$ or $\sigma_C = \sigma_A$ which score B0

Question Number	Scheme	Marks
5.	H_0 : Continuous uniform distribution $[0, 360]$ is a suitable model (for direction of flight). H_1 : Continuous uniform distribution $[0, 360]$ is not a suitable model (for direction of flight).	B1
		Some attempt at $\frac{(\text{Class Width}) \times 450}{360}$ Can be implied by at least one correct E_i — M1
		All expected frequencies are correct. — A1
	At least 3 correct terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. Accept 2 sf accuracy — dM1	
	For applying either $\sum\frac{(O-E)^2}{E}$ or $\sum\frac{O^2}{E} - 450$ — ddM1	
	awrt **14.4** — A1	
	$\nu = 4$ (can be implied) — B1	
	13.277 — B1	
	A correct conclusion in context which is based on their X^2-value and their χ^2-critical value. — A1 ft	
		[9] **9**

Observed	Expected	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
78	90	1.6	67.6
69	85	3.012...	56.011...
51	62.5	2.116	41.616
108	87.5	4.803...	133.302...
144	125	2.888	165.888
	Totals	14.42	464.42

Direction of flight	Expected
$0 \leqslant x < 72$	90
$72 \leqslant x < 140$	85
$140 \leqslant x < 190$	62.5
$190 \leqslant x < 260$	87.5
$260 \leqslant x < 360$	125

$X^2 = \sum\frac{(O-E)^2}{E}$ or $\sum\frac{O^2}{E} - 450$;= awrt 14.4

$\nu = 5 - 1 = 4$

$\chi^2_4(0.01) = 13.277 \Rightarrow$ CR: $X^2 \geqslant 13.277$

[in the CR/significant/Reject H_0]

A continuous uniform distribution is not a suitable model for the direction of flight of honeybees/ Kylie's belief is incorrect.

	Notes
	2nd M1 Dependent on 1st M1 3rd M1 Dependent on both previous M1 marks being scored 3rd A1ft Dependent on all previous M1 marks For a comment in context, follow through their X^2 and with their c.v. Contradictory statements score final A0. E.g. "not significant, Kylie's belief is incorrect."

Question Number	Scheme		Marks
6. (a)	$W = 3X - 4Y$, $X \sim \mathrm{N}(21, 2^2)$, $Y \sim \mathrm{N}(8.5, s^2)$; X, Y are independent.		
	$\{\mathrm{E}(W) = 3\mathrm{E}(X) - 4\mathrm{E}(Y) = 3(21) - 4(8.5)\} \Rightarrow \mathrm{E}(W) = 29$	$\mathrm{E}(W) = 29$ (seen or implied)	B1
	$\mathrm{Var}(W) = 9\,\mathrm{Var}(X) + 16\,\mathrm{Var}(Y)$	$3^2\,\mathrm{Var}(X) + 4^2\,\mathrm{Var}(Y)$	M1
	$\{\mathrm{Var}(W) = 9(4) + 16(\sigma^2)\} \Rightarrow \mathrm{Var}(W) = 36 + 16\sigma^2$	$\mathrm{Var}(W) = 36 + 16\sigma^2$	A1
	$\{\text{So } W \sim \mathrm{N}(29, 36 + 16s^2)\}$		
	$\dfrac{44 - 29}{\sqrt{36 + 16\sigma^2}} = k \ (= 1.2816)$	Standardising $(\pm)$ with their mean and their standard deviation which is in terms of σ^2 and setting equal to k, $\lvert k \rvert > 1$	M1
		± 1.2816 or awrt ± 1.2816	B1
		Correct equation . See notes	A1
	$\sigma^2 = \dfrac{\left(\dfrac{15}{1.2816}\right)^2 - 36}{16} \Rightarrow \sigma = \ldots$	Squaring and rearranging leading to $\sigma = \ldots$	dM1
	$\sigma = 2.51230\ldots = 2.51\,(2\text{dp})$ (= 2.51655... from using 1.28)	awrt **<u>2.51</u>** or awrt **<u>2.52</u>** (only)	A1
			[8]
(b)	$B = 2X + \sum_{i=1}^{3} A_i$, $A \sim \mathrm{N}(28, 5^2)$; X, A_1, A_2 and A_3 are independent.		
	$\mathrm{E}(B) = 2\mathrm{E}(X) + 3\mathrm{E}(A); = 2(21) + 3(28) = 126$	Either $\mathrm{E}(B) = 2\mathrm{E}(X) + 3\mathrm{E}(A)$ or $\mathrm{Var}(B) = 2^2\,\mathrm{Var}(X) + 3\mathrm{Var}(A)$	M1
	$\mathrm{Var}(B) = 2^2\,\mathrm{Var}(X) + 3\mathrm{Var}(A); = 4(4) + 3(5^2) = 91$	At least one of $\mathrm{E}(B) = 126$ or $\mathrm{Var}(B) = 91$	A1
		Both $\mathrm{E}(B) = 126$ and $\mathrm{Var}(B) = 91$	A1
	$\{\text{So } B \sim \mathrm{N}(126, 91)\}$		
	$\left\{\mathrm{P}(B \leqslant 145 \mid B > 120)\right\} = \dfrac{\mathrm{P}(120 < B \leqslant 145)}{\mathrm{P}(B > 120)} =$	A correct conditional probability ratio	M1
	$z_1 = \dfrac{120 - 126}{\sqrt{91}} = -0.62897\ldots$, $z_2 = \dfrac{145 - 126}{\sqrt{91}} = 1.99174\ldots$	Attempt to standardise both 120 and 145 using their $\mathrm{E}(B)$ and their $\mathrm{Var}(B)$	M1
	$= \dfrac{0.7357 - \left(1 - 0.9767\right)}{0.7357}$ (o.e.)	Correct method for finding ***either*** the numerator or the denominator.	dM1
	$= 0.968329\ldots$	awrt **<u>0.968</u>**	A1
	(Calculator gives 0.968449...)		**[7]**
			15

	Notes
(a)	2nd M1 Allow $\dfrac{\pm \text{ their } \mathrm{E}(3X - 4Y)}{\sqrt{\text{their } \mathrm{Var}(3X - 4Y)}} = k$, where $\lvert k \rvert > 1$
	2nd B1 For either -1.2816 or 1.2816
	2nd A1 E.g. Allow $\dfrac{44 - 29}{\sqrt{36 + 16\sigma^2}} = [1.28, 1.29]$, must be compatible signs
	3rd M1 Dependent on the 2nd M1 mark being awarded.
	3rd A1 Dependent on previous A1
(b)	2nd M1 Condone $\mathrm{P}(120 < B < 145)$ but $\mathrm{P}(121 < B < 145)$ is M0
	4th M1 Dependent on the 2nd M1 mark being awarded. (Numerator>denominator is M0).

Question Number	Scheme		Marks
7. (a)	$\left\{\hat{m} = \bar{x} = \dfrac{1152}{8} \Rightarrow\right\}\ \bar{x} = 144$ (grams)	**144**	B1
	$\left\{\hat{\sigma}^2 =\right\}\ s^2 = \dfrac{167218 - 8(144)^2}{8-1} = 190$ (grams)^2	167218 **190**	B1 M1 A1
			[4]
(b)	Contains an unknown parameter / population parameter / μ		B1
			[1]
(c)	$Y = \dfrac{1}{8}\left(\sum_{i=1}^{8} X_i^2 - 8\bar{X}^2\right) = \dfrac{7}{8}S^2$		
	$\left\{\mathrm{E}(Y) = \mathrm{E}\left(\dfrac{7}{8}S^2\right) = \dfrac{7}{8}\mathrm{E}(S^2) \Rightarrow\right\}\ \mathrm{E}(Y) = \dfrac{7}{8}\sigma^2$	$\underline{\frac{7}{8}\sigma^2}$	M1 A1
			[2]
(d)	$\mathrm{bias}(Y) = \dfrac{7}{8}\sigma^2 - \sigma^2; = -\dfrac{1}{8}\sigma^2$	$\underline{-\frac{1}{8}\sigma^2}$ or $\underline{\frac{1}{8}\sigma^2}$	M1 A1
			[2]
			9

	Notes
(a)	2nd B1 For 167218 or $143^2 + 131^2 + 165^2 + 122^2 + 137^2 + 155^2 + 148^2 + 151^2$ (may be implied)
	M1 For use of $\dfrac{'\Sigma x^2' - 8('\Sigma x')}{8-1}$ or $\dfrac{8}{7}\left(\dfrac{'\Sigma x^2'}{8} - ('\bar{x}')^2\right)$ where $\Sigma x^2 \neq 20736$
(c)	M1 For $k\sigma^2$, where $0 < k < 2, k \neq 1$
(d)	M1 For their $\pm\left(\mathrm{E}(Y) - \sigma^2\right)$, where their $\mathrm{E}(Y) \neq \sigma^2$.

Question Number	Scheme	Marks
8.	Let X = score on a die, $X \sim \text{Bin}\left(30, \frac{1}{6}\right)$, $\text{E}(X) = 5$, $\text{Var}(X) = \frac{25}{6}$	
(a)	$[\bar{X} \sim]\text{N}\left(5, \frac{1}{12}\right)$	B1dB1B1 **[3]**
(b)	CR: $\frac{\bar{X} - 5}{\sqrt{\frac{1}{12}}} \leqslant -1.96$ or $\frac{\bar{X} - 5}{\sqrt{\frac{1}{12}}} \geqslant 1.96$ 1.96 or -1.96	M1 B1
	CR: $\bar{X} \leqslant 4.434196...$ or $\bar{X} \geqslant 5.565803...$	A1 A1 **[4]** **7**
	Notes	
(a)	1st B1 Normal or N 2nd B1 dependent on 1st B1 for mean of 5 3rd B1 $\text{Var}(\bar{X}) = \frac{1}{12}$ oe	
(b)	M1 for an attempt to standardise using their $\text{E}(\bar{X})$ and their $\text{Var}(\bar{X})$ and setting $\leq -z$ **or** $\geq z$ $(z > 1)$ 1st A1 for at least one of either $\bar{X} \leqslant$ awrt 4.43 **or** $\bar{X} \geqslant$ awrt 5.57 or $\bar{X} \geqslant$ trunc. 5.56 2nd A1 both $\bar{X} \leqslant$ awrt 4.43 **and** either $\bar{X} \geqslant$ awrt 5.57 or $\bar{X} \geqslant$ trunc 5.56	

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom