edexcel

Mark Scheme (Results)

June 2012

GCSE Statistics
Paper: 5ST1H/01

Edexcel and BTEC Qualifications
Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please call our GCE line on 0844 576 0025, our GCSE team on 0844 576 0027, or visit our qualifications website at www.edexcel.com. For information about our BTEC qualifications, please call 0844 576 0026, or visit our website at www.btec.co.uk.

If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

Ask The Expert can be accessed online at the following link:

http://www.edexcel.com/Aboutus/contact-us/

Alternatively, you can speak directly to the subject team at Pearson about Edexcel qualifications. Their contact details can be found on this link: www.edexcel.com/teachingservices

Pearson: helping people progress, everywhere
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 raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

June 2012
Publications Code UG033134

NOTES ON MARKING PRINCIPLES

1 **Mark Schemes**

- These should be applied positively. Candidates should all receive the same treatment. They should be rewarded for what they haveshown they can do rather than penalised for omissions.

2 **Types of mark**

- M marks: method marks
- A marks: accuracy marks Note: you cannot give an A mark if you have given M0
- B marks: unconditional accuracy marks (independent of M marks)

3 **Abbreviations**

- cao – correct answer only
- isw – ignore subsequent working
- oe – or equivalent (and appropriate)
- indep – independent
- QWC – quality of written communication
- ft – follow through
- SC: special case
- dep – dependent
- awrt – anything which rounds to
- ( ) – brackets round words mean these are not essential

4 **No working**

- If no working is shown then correct answers normally score full marks
- If no working is shown then incorrect (even though nearly correct) answers score no marks.

5 **With working**

- If there is a wrong answer indicated on the answer line always check the working in the body of the script (and on any diagrams), and award any marks appropriate from the mark scheme.
- If working is crossed out and still legible, then it should be given any appropriate marks, as long as it has not been replaced by alternative work.
- If it is clear from the working that the "correct" answer has been obtained from incorrect working, award 0 marks. Send the response to review, and discuss each of these situations with your Team Leader.
- If there is no answer on the answer line then check the working for an obvious answer.
- Any case of suspected misread loses A (and B) marks on that part, but can gain the M marks. Discuss each of these situations with your Team Leader. If there is a choice of methods shown, then no marks should be awarded, unless the answer on the answer line makes clear the method that has been used.

6 **Follow through marks**

- Follow through marks which involve a single stage calculation can be awarded without working since you can check the answer yourself, but if ambiguous do not award.
- Follow through marks which involve more than one stage of calculation can only be awarded on sight of the relevant working, even if it appears obvious that there is only one way you could get the answer given.

7 **Ignoring subsequent work**

- It is appropriate to ignore subsequent work when the additional work does not change the answer in a way that is inappropriate for the question: e.g. incorrect cancelling of a fraction that would otherwise be correct
- It is not appropriate to ignore subsequent work when the additional work essentially makes the answer incorrect e.g. algebra.
- Transcription errors occur when candidates present a correct answer in working, and write it incorrectly on the answer line; mark the correct answer.

8 **Probability**

- Probability answers must be given a fractions, percentages or decimals. If a candidate gives a decimal equivalent to a probability, this should be written to at least 2 decimal places (unless tenths), unless it states otherwise on the mark scheme.
- Incorrect notation should lose the accuracy marks, but be awarded any implied method marks.
- If a probability answer is given on the answer line using both incorrect and correct notation, award the marks.
- If a probability fraction is given then cancelled incorrectly, ignore the incorrectly cancelled answer.

9 **Linear equations**

- Full marks can be gained if the solution alone is given on the answer line, or otherwise unambiguously indicated in working (without contradiction elsewhere). Where the correct solution only is shown substituted, but not identified as the solution, the accuracy mark is lost but any method marks can be awarded.

10 **Parts of questions**

- Unless allowed by the mark scheme, the marks allocated to one part of the question CANNOT be awarded in another.

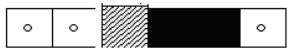
11 **Range of answers**

- Unless otherwise stated, when an answer is given in a range (e.g. 3.5 – 4.2) then this is inclusive of the end points, and includes all the numbers in between.

12 **Quality of Written Communication**

- This is denoted by an asterisk near the question number/part (*). Mark schemes will indicate within the table how marks are to beallocated. In this subject we need to see that correct statistical terms are used.

Guidance on the use of codes within this mark scheme
M1 – method mark A1 – accuracy mark B1 – Working mark C1 – communication mark QWC – quality of written communication oe – or equivalent awrt – anything which rounds to cao – correct answer only ft – follow through sc – special case dep – dependent (on a previous mark or conclusion) indep – independent isw – ignore subsequent working

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
1	(a)(i)		Illness (not medical or dental appt.)	2	B1 Accept ‘illness’.
	(ii)		Other unauthorised circumstances		B1 Accept ‘other’.
	(b)		Excluded, no alternative provision	1	B1 Accept ‘excluded’. May be seen in a sentence.
2	(a)			2	B2 if all 6 correct (B1 for 5 correct)
	(b)		Top right of grid	2	B1 Accept equivalent descriptions, including reference to individual square**s** in this area. (...but not just a single square) e.g. “Square F1” is B0. “Around square F1” is B1 “along E” is B0. “Top of E” is B1 Accept North East.
			Squares are shaded darkest in this area. oe		B1 Accept sensible equivalent wording. (Reference to individual numbers is B0) e.g. ... most black/solid squares is B1 ... most 16-24 squares is B1 BUT ... likely to be where a building stood is B0

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
3	(a)	ANY TWO FROM THREE: 1. Cheaper 2. Less time/quicker 3. Less data/easier (to handle)		2	B2 for two correct (B1 B0 for one correct) Accept equivalent statements. Accept two statements in one answer. Accept converses if clearly refer to 'census'. (One comment only from each type.) (Do not allow contradictory comments.)
	(b)	Electoral **roll** or electoral **register** OR **A list of** council tax payers/residents (register or database are equivalent to list)		1	B1 A suitable list of the population is required. (incomplete lists: e.g. telephone directory is B0; all council tax payers is B0)
	(c)	ANY TWO OF: • This is not a good sample. • This is biased. • Not all residents have an equal chance of being selected. • Only asks North Street residents. • Residents elsewhere cannot give opinions. • Residents in one street may have similar interests/views. • North Street may not be representative. • (Sample) too small		2	B2 Two correct statements (B1 for one correct statement). Allow sensible equivalent wording suggesting bias, restricted opinions, or too small sample. Ignore excess reasons if not contradictory.
	(d)(i)	EITHER This is biased/leading. OR This is trying to persuade you to agree oe		2	B1
	(ii)	EITHER The boxes overlap/You do not know which box for £2 (or £1, or £0) OR It doesn't say how often (weekly/monthly etc) OR No option for not wanting one.			B1 Accept e.g. it should say £1 to £1.99 e.g. it should say how much **per week/month** etc

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
4	(a)		26, 53, 80, 97, 100	1	B1
	(b)	Correct plotting of points and joining with curve or straight lines	Correct Horizontal Plots Correct Vertical Plots	3	B1 B1 ft (½ square tolerance for these two marks) **SC** If B0 B0 then five correct points is B1
			Join with straight lines or curve.		B1ft (Must be increasing) (Ignore any line to left of 20 – i.e. (16, 0) not needed)
	(c)	line from 50 on 'vertical' scale		2	M1 (implied by A1ft within tolerance) (Allow use of their $\Sigma f \div 2$)
			29.5		A1 ft (Accept value in range **29 up to but <u>less than</u> 30**, OR ft value from a line/curve with positive gradient, ±1 small square)
	(d)		The age (at which women have their first child) has gone up/increased. OR Women (having their first child) **in 2008** are **older**.	1	B1 ft from (c) (**If no answer to (c)**, then B0) Accept equivalent clear **comparison** of 2008 and 1990. e.g. "is/was lower" is B0 but "is/was lower in 1990" is B1 (Ignore any figures. e.g. gone up by x years) Must be comparison, not just stating values.

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
5	(a)		8 9 1 4 8 3 4 5 7 7 9 0 1 3 5 5	2	B2 All correct (B1 One error or omission) (Note **misplaced** leaf is **one** error not two) NB: Ignore '6' in median position (the child added in part (d)) **SC**: Unordered (with all leaves) is B1
	(b)		16.6	1	B1
	(c)	$\frac{262.9}{16}$	16.4(3125)	2	M1 for Σx divided by 16 A1 awrt 16.4 (allow $16\frac{69}{160}$)
	(d)(i)		It will stay the same.	2	B1 Allow clear equivalent wording **SC**: "It will become 16.6" is B1 (BUT 16.6 alone is B0)
	(ii)		It will go up/increase (slightly) OR It will change to **16.44**(1176...) OR It will stay the same **<u>to 1 dp</u>**		B1ft Correct interpretation from their (c) e.g. "still 16.4" is B0 but "still 16.4 to 1dp" is B1
	(e)		$\frac{5}{17}$	2	M1 for seeing 5 or $\frac{5}{n}$ A1 cao (allow 0.29 or 29% awrt) **SC**: 0.3125 ($=\frac{5}{16}$) is B1

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
6	(a)	'Males are unemployed for longer/more months than females.' (or converse) OR 'males and females are unemployed for similar time'		1	B1 Any hypothesis which relates to **length of unemployment** of **both** males and females. (Reference to numbers rather than % may be acceptable in this part only.) (Do NOT accept questions.)
	(b)	Lines at 44 then 64 then 81 then 91 then 95, and correct shading. (Tolerance of ½ line vertically)		3	M1 for attempt at drawing a composite (i.e. stacked) chart, blocks in correct order. A2 for getting all correct OR A1 for getting lower 2 blocks correct (44,64) or all correct lines but with incorrect/no shading. **SC**: If M0, all correct lines within tolerance is B2 (ignore shading).
	(c)	CONCLUSION: My hypothesis is/is not correct. (May be a re-statement of hypothesis in (a) **OR** answering their 'question' from (a))		2	B1ft **Must ft correctly from part (a).** Note: If hypothesis/question in (a) relates to numbers unemployed, then first B1 in (c) can **only** be awarded for "cannot say if correct, as we only have percentages".
		REASON: e.g. Higher % of females unemployed for shorter periods. e.g. 52% males and 64% females under 6 months unemployed. e.g. 14% males but only 5% females over 36 months unemployed. etc			B1ft Allow follow through from (b). A correct comment **comparing** a **number of months** of unemployment Candidates must **NOT** refer to **numbers of males/females**. This question is about %

5ST1H_01

Question	Working	Answer	Mark	Notes
7				
(a)(i)		**Qual**itative	1	B1
(ii)		**Quant**itative	1	B1
(iii)		**Quant**itative	1	B1
(b)		Speed (of serve)	1	B1
(c)				B1 for ranks correct
	$\Sigma d^2 = 128$ (Reverse: $\Sigma d^2 = 202$)			M1 for Σd^2 attempted for their ranks (must use **ranks**)
	$r_s = 1 - \frac{6\times'128'}{10\times 99}$ $\left(\text{Reverse}: \ 1-\frac{6\times'202'}{10\times 99}\right)$			M1(**dep** on 1st M1) for correct substitution into formula (including 1-...) with their Σd^2
		= **±0.22(4...)**	4	A1 awrt ±0.22
(d)			2	B1 Description consistent with (c) (Only follow through if $\lvert r_s\rvert<1$)
				B1 cao (Interpretation) Answer in context. **Must not be inconsistent with description. Must not be inconsistent with their ranking.** ($\lvert r_s\rvert<1$) (If B0 for description it must be consistent with answer to (c))

rank	d^2	*rank'*	d^2
1	0	*10*	*81*
7	25	*4*	*4*
4	1	*7*	*16*
5	1	*6*	*4*
10	25	*1*	*16*
9	9	*2*	*16*
2	25	*9*	*4*
3	25	*8*	*0*
8	1	*3*	*36*
6	16	*5*	*25*

Description:	This is (weak) **positive correlation** (or negative if (c) is <0)		This is **no/little/weak correlation**
Interpretation:	(There is slight evidence to say) the faster the serve the higher the position in world tennis (rank)	OR	There is no (clear evidence of a) relationship between serve and position in world tennis (rank).

5ST1H_01

<table>
<tr><th colspan="2">Question</th><th>Working</th><th>Answer</th><th>Mark</th><th>Notes</th></tr>
<tr><td>8</td><td>(a)</td><td>ANY TWO OF:
1. Time Consuming.
2. More expensive.
3. Lots of data to handle
OR harder/more work than a sample.
4. Difficult to be sure the whole population is used.</td><td></td><td>2</td><td>B2 for two correct
(B1 B0 for one correct)
May give two reasons in one answer space.
Do not allow same reason type twice. Do not accept contradictory statements.</td></tr>
<tr><td></td><td>*(b)</td><td colspan="2">
<table>
<tr><th colspan="2">Aspect</th><th>Answer</th><th>SC1</th><th>SC2</th></tr>
<tr><td>1</td><td>Name</td><td>Use Stratified sampling</td><td>Systematic sampling</td><td>Random sampling</td></tr>
<tr><td>2</td><td>Number-ing</td><td>Number students</td><td>Number students</td><td>Number students</td></tr>
<tr><td>3</td><td>Random-ness</td><td>Select by random sampling</td><td>With a random starting point (between 0–7 (or 1- 8) select every 8th student.)</td><td>Create 100 random numbers (and use students with these numbers as your sample)</td></tr>
<tr><td>4</td><td>Proport-ion</td><td>Numbers in proportion with the year group size
OR
Same percentage from each year group
OR
18 students from Y7
(<u>or</u> 18 of Y8, 19 of Y9,
12 of Y10 or Y11,
10 of Y12, or 11of Y13)
(i.e. calculation for at least one year group)</td><td></td><td></td></tr>
</table>
Note: Significance of <u>wording in bold</u> is shown in <u>5th B1 mark</u>
</td><td>5</td><td>B1 (aspect 1) Name
B1 (aspect 2) Numbering
B1 (aspect 3)
Idea of randomness
('Words' in brackets not needed for this B1)
B1 (aspect 4)
Apply Special Cases SC1 & SC2 if not describing proportion/strata approach.
B1 Work which uses the correct statistical 'wording' shown in bold (including that in brackets) <u>in any 3 aspects</u> for <u>their method</u>.</td></tr>
</table>

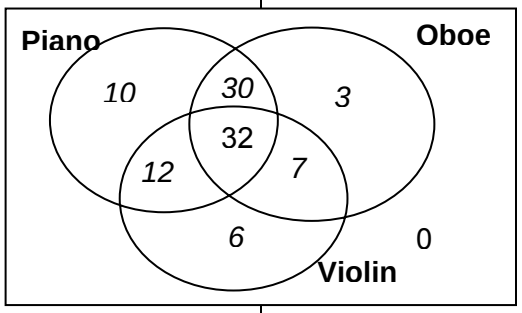
5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
9	(a)		Class Widths: (4), 5, 10, 10, 10, 10, 20 FD: (3), 3, 2, 1.8, 1.7, 1.1, 0.35	3	B1 correct widths B2 all fd correct (B1 if at least three fd values correct OR B1ft if **ALL** fd follow through from their widths)
	(b)		Correct histogram	2	B2ft All correct from their frequency **density** (½ small square tolerance) (B1ft if only one incorrect block)
	(c)		This can be seen by the **Positive Skew** of the diagram. OR Some sensible statement about the shape of the histogram or frequency densities. The probability of a young driver having an accident is greater. (OR converse with older driver) OR Premiums are related to risk.	2	B1 (Reason) Allow "higher bars (or higher fd) for younger drivers" OR "bars are decreasing" B1 (Contextual interpretation) Allow "younger drivers (or "they") have more accidents"

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
10	(a)(i)	$\frac{9.4+7.0+6.2+8.0}{4}$ $\frac{7.0+6.2+8.0+8.8}{4}$ $\frac{6.2+8.0+8.8+7.0}{4}$	7.65, 7.5, 7.5 (millions)	4	M1 for attempt to add 4 numbers from table and divide by 4 (Implied by one correct if no working) A2 for 3 correct (any order) (A1 for 2 correct seen)
	(ii)		Plot 3 points - Mid way 2008 Q4/Q1, 2009 Q1/Q2, 2009 Q2/Q3		B1ft for plotting all three correctly (ie ½ square tolerance)
	(b)		Trend line	1	B1 Appropriate trend line
	(c)		Downward / decreasing / falling OR e.g. there are fewer visitors as time goes on	1	B1 Condone 'negative (trend)' but 'negative correlation' alone is B0 Direct comparison of dates is B0
	(d)	$\pm\frac{'1.1'+'1.5'+'1.5'}{3}=\frac{4.1}{3}$	= 1.3...	2	M1 for adding 3 attempts at seasonal variation (consistent quarter) **and** dividing by 3 A1ft (positive only) (Accept **1.2** to **1.4**, **or** ft from their line)
	(e)	Finding value ('6.8') in range 6.5 to 7.0 **OR** finding value from their extended trend line at 2010/Q3 '6.8' + '1.3...'	= 8.1...	3	M1ft '6.8' (in range 6.5 to 7.0) OR for their trend line value M1d ft (**Dep** on 1st M1) for their '6.8' +|their (d)| (Only if 1 < |their (d)| < 2) A1 Must see some working. (Accept answer in range **7.7** to **8.4**)

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
11	(a)		4.2 hrs, $4\frac{1}{5}$ hrs, $4\frac{2}{10}$ hrs OR 4 hours 12 minutes	1	B1 (Allow 4h 12min, BUT 4.12 or 4.12min are B0)
	(b)		5 hours (5 hrs 0 min))	1	B1
	(c)(i)		X is an outlier (OR anomaly OR extreme/rogue value)	4	B1 Condone poor spelling but not 'outliner'
	(ii)	(IQR =) 6 - 5	The IQR is 1 (hour).		M1 for IQR=1, OR 6-5(=1)
		1.5 × 1 = 1.5 hours 6 + 1.5 = 7.5 hrs	Upper Outliers start at 6 + 1.5 (hours)		M1 7 h 30 min or 7.5 (or 6 + 1.5) (Their Q3 + **1.5x** their IQR)
		Outlier if > 7.5 (or hours and minutes)	Values above 7.5 are outliers OR **8.9** (or **X**)> **7.5 h**ours OR X is more than 1.5 IQR (or 1.5 hrs) above upper quartile (or 6h)		A1 dep on both M1 There must be reference to their '7.5' being an outlier boundary **OR** a clear comparison of X or 8.9 with 7.5
	(d)		Four <u>comparisons</u> from: • Males have greater **Interquartile range** (or **IQR**) • Males have a greater **range** • Males have a lower **median** • Males distribution is **symmetrical** (or **no skew** or **no outlier**) but Females distribution is **positive skew** (OR both are roughly **symmetrical**) • Contextual interpretation of one of the above	4	B1 B1 B1 B1 Max one mark from each of the five options. Ignore excess statements if not contradictory. Allow equivalent converse statements about females. Comparison of individual values other than median is B0 The words in bold must be used in those comparisons. (Condone poor spelling but 'medium' is B0)

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
12	(a)(i)		$\frac{280}{20}$ or 280÷20 OR 14x20=280	3	B1
	(ii)	$\sqrt{\frac{4220}{20}-14^2}$	$\sqrt{\frac{4220}{20}-14^2}$ = √15 = 3.87… = 3.9		M1 for using correct formula for sd, including square root. **NB**: $\sqrt{\frac{4220}{280}} = 3.88\ (=3.9)$ Scores M0A0 A1 evaluated to 3.87 or better **OR** √15 leading to 3.9
	(b)	$\frac{12-14}{3.9}$	(= 0.516…) = - 0.5 awrt	3	M1 for sight or use of $\pm\left(\frac{x-\text{mean}}{\text{sd}}\right)$ M1 for $+\left(\frac{12-14}{3.9}\right)$ (with 3.9 or better for sd) A1 for awrt -0.5 only **SC:** No working with awrt **+**0.51 or **+**0.52 is B1
	(c)		Melvin did better in the figure skating. His **standardised score** is **higher** in figure skating. (OR converse)	2	**Both these marks are dependent upon having an answer to (b).** B1ft for correct conclusion for their value in (b) B1ft for correct comparison of values. Must follow for their (b) in range -4 to +4 (Less negative is B1 BUT Closer to zero/mean is B0)

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
13	(a)		COMPARISON: There are more (people living in the Lake District) REASON: The pie chart of the people living in the Lake District has the largest area. (accept 'is bigger') 2		**For both marks: only allow converse if they mention 'Snowdonia'.** B1 Any comment implying more people: e.g. 2.25x (or 2x) as many people B1 Any comment implying the chart is bigger: e.g. 2.25x as big (or 2x as big) e.g. bigger radius/diameter
	(b)		40 – 59 (years)	1	B1
	(c)		COMPARISON: There are more (40 -59 year olds living in the Lake District.) REASON: The **area** of the sector **is greater** (in the pie chart for the Lake District.) 2		**For both marks: only allow converse if they mention 'Snowdonia'.** B1 B1 Need to see the word 'area', BUT Accept: It is the **same proportion/angle** of a **larger population/pie chart** IF **'area'** is mentioned in (a)

5ST1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
14	(a)		30,12 and 7 in correct place 10 and 6 in correct place 3 in correct place	3	B1 (30, 12, 7) B1 (10, 6) B1ft (3) Award final B1ft if their '10'+'6'+'3' is 19
					
	(b)	$\frac{32+'30'}{100}$	$\frac{62}{100}=\frac{31}{50}$ oe	1	B1ft (follow through their 30, but answer must be a probability)
	(c)	$\frac{'12'+32}{84}$	$\frac{44}{84}=\frac{11}{21}$ oe (0.52 or better)	2	M1 for $\frac{n}{84}$ (must have $0 < n < 84$) (This can be implied by correct answer, or by correct follow through answer) A1 ft (follow through their 12, even with no working)

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com

Order Code UG033134 June 2012

For more information on Edexcel qualifications, please visit
www.edexcel.com/quals

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning