Paper 1 Higher tier mark scheme

Question number	Answer	Additional guidance	Mark
1	B1 for e.g. median greater after revision B1 for e.g. IQR/range smaller after revision B1 for e.g. negative skew after revision, symmetrical before revision B1 for e.g. • They did better after revision • They were more consistent after revision	B1 for a correct statistical statement comparing the medians B1 for a correct comparison of the IQRs or ranges B1 for a correct comparison of the skews B1 for a correct contextual interpretation comparing medians or IQR/ranges	**(4)**
2(a)	B1B1 for two correct reasons • e.g. data given in different formats • e.g. remove extraneous symbols • e.g. remove anomalies/outliers • e.g. data given in wrong order	B1 for each correct reason for the need to clean data on the database prior to processing it	**(2)**
2(b)	B1B1 for two correct aspects • e.g. large sample size increases reliability • e.g. issues due to how the data collection is carried out may decrease reliability (recorded by students and not by Kerry/students may type in information wrong) • e.g. for an example of factors that might not be consistent in the data collection which may decrease reliability (last message might not represent all text messages sent) • e.g. non-response decreases reliability	B1B1 for two correct comparison assessing the reliability of the conclusions drawn	**(2)**

<table>
<tr><th>Question number</th><th>Answer</th><th>Additional guidance</th><th>Mark</th></tr>
<tr><td>3(a)</td><td>B1 3 bedrooms has the highest frequency (for any individual number of bedrooms)</td><td>Allow equivalent statistical reasoning based on the table indicating why 3 is the mode.</td><td>(1)</td></tr>
<tr><td>3(b)</td><td>M1 $\frac{140}{1200}\times 60$ oe

A1A1<table><tr><td>Bedrooms</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5+</td></tr><tr><td>Houses in sample</td><td>7</td><td>15</td><td>21</td><td>12</td><td>5</td></tr></table></td><td>Accept a correct equivalent calculation shown for any one class
M1 implied by one correct answer OR an indication they need 1 in 20

1st A1 for any one value correct
2nd A1 for all correct</td><td>(3)</td></tr>
<tr><td>3(c)</td><td>B1 Use a sampling frame for each strata

B1 Select houses randomly or generate random numbers

B1 For an aspect of detail</td><td>Each category/strata to be considered separately

Samples have to be random

e.g. How the random numbers are obtained and used</td><td>(3)</td></tr>
<tr><td>4(a)</td><td>M1 $\frac{4}{30}\times 180$
A1 $= 24$</td><td>Accept any equivalent calculation</td><td>(2)</td></tr>
<tr><td>4(b)</td><td>B1 e.g. ‘use the mean of the sample of an estimate’</td><td>B1 for a correct explanation showing understanding that the mean of the sample represents the mean of the population</td><td>(1)</td></tr>
</table>

Question number	Answer	Additional guidance	Mark
5	M1A1A1 for appropriate calculations to compare 1996 with 2006 or 1996 with 2016 e.g. **Comparing index numbers** M1 $\frac{23}{20}\times 100$ or $\frac{54}{20}\times 100$ A1 115 A1 270 OR **Comparing prices** M1 $20\times\frac{129}{100}$ or $20\times\frac{172}{100}$ A1 25.8 A1 34.4 B1 for e.g. from 1996 to 2006 the change/increase in price was less than the RPI B1 for e.g. from 1996 to 2016 the change/increase in price was higher than the RPI	M1 for a calculation that could be used to compare 1996 with 2006 or 2016 A1 for one correct value A1 for second correct value that allows a comparison to be made B1 for a correct contextual statement for 2006 or 2016 B1 for a correct contextual statement for 2006 and 2016	**(5)**

Question number	Answer	Additional guidance	Mark
6(a)	B2 for weighted mean (or Paul) AND reference to there being different proportions/percentages of people earning each weekly amount OR B1 for weighted mean (or Paul) with attempt at reason	B2 for a complete assessment of the appropriate choice with reason OR B1 for an incomplete assessment of the appropriate choice	**(2)**
6(b)	M1 for $0.2\times260 + 0.35\times370 + 0.45\times510$ A1 for 411	M1 for a correct method to find the weighted mean A1 for the correct answer	**(2)**
7	B1 for the questions are personal/people may be embarrassed AND B2 for a correct comment assessing the appropriateness of the interview and a reason e.g. • appropriate and e.g. an interview will have a good response rate (or higher response rate than e.g. postal survey) • appropriate and e.g. the interviewer could be trained to put people at their ease when answering the personal questions • not appropriate and e.g. people may not feel comfortable talking about their health/fitness with the interviewer OR (if B2 not earned) B1 for a correct commenting relating to the appropriateness of the interview without a decision	B1 for assessing appropriateness of questions B2 for a correct comment assessing the appropriateness of the interview and a reason OR (if B2 not earned) B1 for an incomplete assessment of the appropriateness of the interview	**(3)**

Question number	Answer	Additional guidance	Mark
8(a)	M1 for either 111.2/112.7×100 (=98.7) OR 113.1/111.2×100 (=101.7) A1 for 98.1 and 10.7	M1 for correct calculation of chain base index number. May be implied by one correct answer A1 both correct	**(2)**
8(b)(i)	M1ft for $\sqrt[4]{}$(102.5×100.8×'98.7'×'101.7') A1ft for100.9	M1ft for correct calculation of the geometric mean of the four chain base index numbers A1ft correct answer ft their answers in (a)	**(2)**
8(b)(ii)	B1ft for (average) rate of 'increase' ... B1ft ... is '0.9'% per month	B1ft for correct contextual interpretation as rate of increasing B1ft for complete correct contextual interpretation of their value for geometric mean	**(2)**
9	B2 for not reliable with a correct reason e.g. samples too small or time interval between samples too long (as population may have changed between samples)	B2 for a correct comment assessing the appropriateness of the conclusion OR if B2 not earned B1 for an incomplete attempt to assess the appropriateness of the conclusion	**(2)**

Question number	Answer	Additional guidance	Mark
10(a)	B1 for a correct statement identifying any seasonality, e.g. the greatest/least values are in quarter 3/ quarter 1 B1 for a correct interpretation in context for the identified seasonality, e.g. more/less overseas visitors in summer/winter	B2 for a correct statement identifying and interpreting an example of seasonality (B1 for one of these features)	**(2)**
10(b)	M1 Correct horizontal plotting, first between Q2/Q3 of 2012 A1 All points correct B1 Straight trend line through moving averages within tolerance	M1 for recognising correct horizontal position for plotting moving averages (at least five correct) A1 for accurately plotting all moving averages (Allow within one small square tolerance) B1 Their line should extend horizontally at least from 2012 Q3 to 2013 Q4 and be vertically within one square of 7.8 at 2012 Q3 and one square of 8.4 at 2013 Q4	**(3)**
10(c)	B1 Upward/rising trend	Accept equivalent wording demonstrating statistical reasoning	**(1)**
10(d)	B1ft trend line value 8.9 (±0.1) M1 for $8.9 - \frac{(7.55-6.25)+(8-6.3)+(8.45-6.8)}{3}$ A1 for 7.35	B1ft for value which follows from their line at 2015 Q1 M1 for complete method to find seasonal variation for their trend line A1 for answer in range [7.2−7.5]	**(3)**
10(e)	B1 for e.g. the pattern in the data repeats after four quarters	B1 for a correct statement assessing the appropriateness of using 4-point moving averages	**(1)**

Question number	Answer	Additional guidance	Mark
11(a)	M1 difference in ranks: 0, 1, −2, −1, −3, 2, 2, 1, 0 M1 $(r_s =)\ 1-\frac{6\times 24}{9\times(9^2-1)}$ A1 0.8 B1ft Positive (rank) correlation B1ft Judges were in agreement with the public	M1 for difference in ranks (condone one slip and allow ±). Can be implied by $\Sigma d^2 = 24$ M1 for demonstrating correct use of Spearman's formula A1 cao B1ft for statistical interpretation of their '0.8' B1ft for correct contextual conclusion from their '0.8'	**(5)**
11(b)(i) **11(b)(ii)**	B1 **both negative** values identified for the graph B1 Spearman's = −0.9 AND pmcc = −0.7 B1 pmcc is less strong correlation as it measures closeness to a **linear** model	B1 for recognising graph will give negative correlation B1 cao, for recognising pmcc calculation will be closer to zero when correlation is non-linear B1 for equivalent statistical reasoning that pmcc will be closer to 0 as graph does not suggest a straight line.	**(3)**

Question number	Answer	Additional guidance	Mark
12(a)	B1 Graph supports the hypothesis + reason B1 Scatter shows **negative correlation**	B1 for conclusion supported by sensible reason B1 for statistical reasoning using words in bold	**(2)**
12(b)	M1 e.g. $83.5 - 0.7 \times 4 = 80.7$ or $83.5 - 0.7 \times 8 = 77.9$ or sensible straight line with correct gradient A1 correct straight line within tolerance	M1 for one pair of coordinates correctly identified, or for a sensible attempt at straight line with correct gradient. A1 Their line should extend horizontally at least from 4.2 to 7 and (if extended) be vertically within one square of 80.7 at $x = 4$ and one square of 77.9 at $x = 8$	**(2)**
12(c)	B1 79.6 from graph	B1 for answer between 79.5 and 79.7 (may use equation OR vertical line drawn from $x = 5.6$)	**(1)**
12(d)	B1 e.g. change in life expectancy as unemployment increases. B1 **0.7 years fall** in life expectancy (per 1% increase in unemployment)	B1 for recognising in context that gradient indicates a **rate**. Accept equivalent wording. No figures needed for 1st B1 B1 for interpreting in context the value. Mark may be gained for correct equivalent figures used within their comment. (e.g. 1.4% more unemployment results in 1 year reduction in life expectancy)	**(2)**
12(e)	B1 B1 for any two comments from • Involves extrapolation/8% is outside of range • It is a for a different region so may be affected by other factors • Correlation does not look very strong	Accept equivalent statistical reasoning. B1 for each correct point but allow each bullet point once only.	**(2)**

Question number	Answer	Additional guidance	Mark
13(a)	B1 for e.g. this is a sensitive question or people may not want to answer it otherwise	B1 for a correct response referring to the sensitivity of the question	**(1)**
13(b)	e.g. M1 for 0.5 × (743 + 679) (= 711) M1 for $\frac{743-"711"}{"711"}$ A1 for 0.045(007…)	M1 for method to estimate the number of people who answered yes because they got Heads M1 for method to estimate the proportion of people who have downloaded illegally A1 for a correct proportion, e.g. 0.045 or 4.5% or better	**(3)**
13(c)	B1 for not appropriate B1 for a correct reason, e.g. the town may not be representative of the UK B1 for a different correct reason, e.g. the telephone directory may not include everyone in the town	B1 B1 B1 for assessing the appropriateness of the statistical methodology with correct reasons	**(3)**

Question number	Answer	Additional guidance	Mark
14(a)(i)	B1 for binomial (distribution)	B1 for correctly identifying the binomial distribution	**(2)**
14(a)(ii)	B1 for a correct property, e.g. independent trials or the 6 people are not related OR only two possible outcomes, i.e. success/failure OR fixed number of trials, i.e. $n = 6$, etc	B1 for a correct property needed for the binomial distribution	
14(b)	M1 for $1 - 0.04$ ($= 0.96$) M1 for $6 \times 0.04 \times 0.96^5$ A1 for 0.196	M1 for finding the probability of not AO M1 for a complete method A1 for answer in range 0.19 – 0.20	**(3)**
14(c)	M1 for $1 - 0.96n$, for any $n \geq 2$ A1 for ($n = 17$) B1 for a correct conclusion, e.g. the smallest value of n is 17	M1 for a correct method to work out at least one probability A1 for the correct answer (17) associated with the correct probability (0.5004…) B1 for a correct interpretation of the answer (17) in the context of the problem	**(3)**